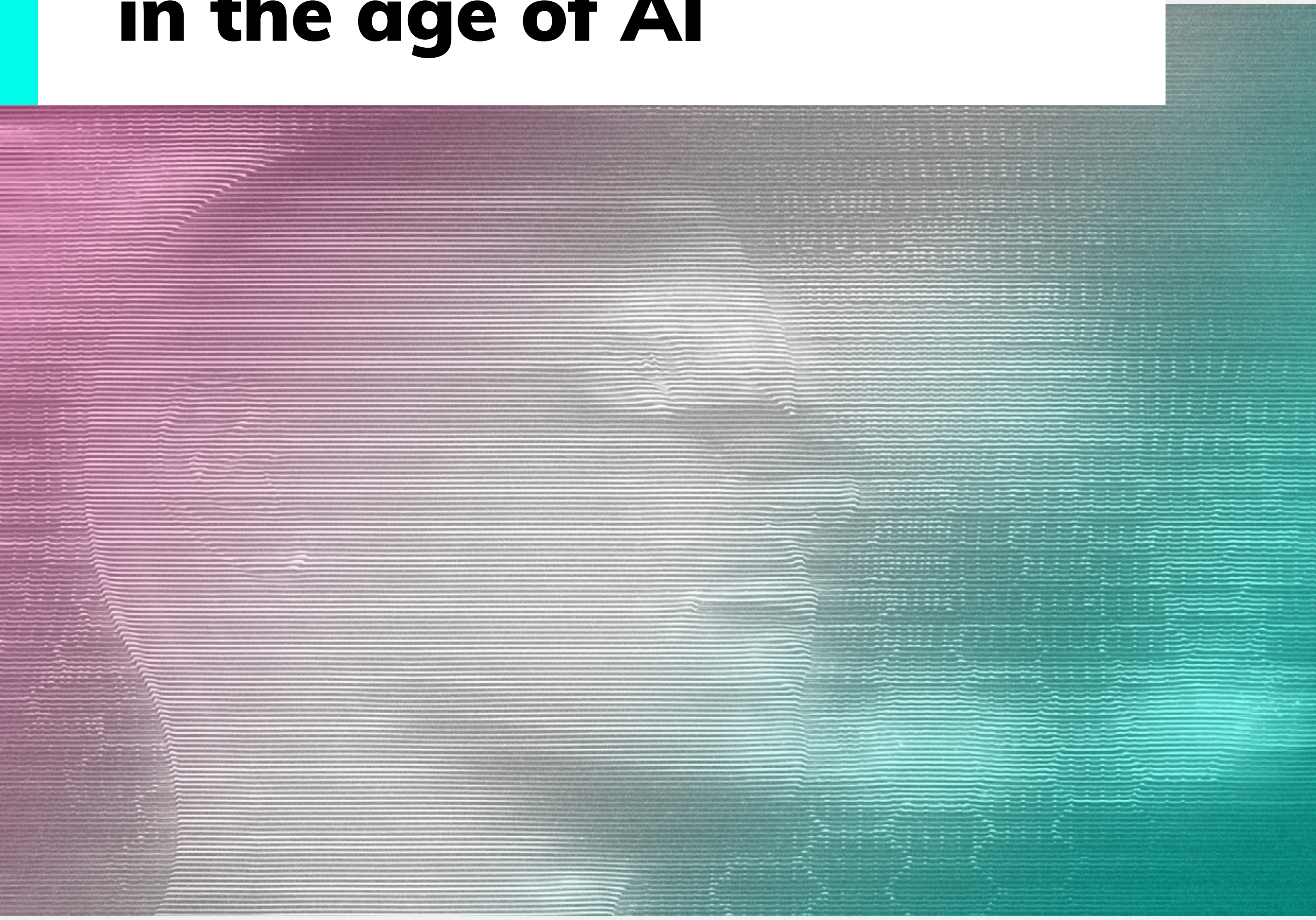
APRIL 2024

Full Fact Report 2024

Trust and truth in the age of AI



FULL FACT

Full Fact Report 2024

Trust and truth
in the age of AI

Contents

About this report

Full Fact fights bad information[1]. We do this in four main ways. We fact check claims made by politicians and public institutions, in the press and online. We then follow up on these, to stop and reduce the spread of specific claims. We campaign for systems changes to help make bad information rarer and less harmful, and advocate for higher standards in public debate.

This report explores how generative AI has presented challenging new risks and opportunities to those tackling misinformation and disinformation, and then looks at progress on themes raised in previous reports.

This report follows on from our 2023 report *Informed citizens: Addressing bad information in a healthy democracy* and our 2022 report *Tackling online misinformation in an open society—what law and regulation should do*, along with earlier reports. This is the fifth report that we have been able to produce thanks to the support of the Nuffield Foundation.

The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in education, welfare, and justice. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics, the Ada Lovelace Institute and the Nuffield Family Justice Observatory. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation.

This report was written by staff at Full Fact and the contents are the responsibility of the Chief Executive. They may or may not reflect the views of members of Full Fact's cross-party Board of Trustees.

We would like to extend our warmest thanks to Maeve Walsh, Raquel Vazquez Llorente, Declan Shaw, Gavin Freeguard and Mark Franks for their comments on earlier versions of this report.

In addition, we thank our other supporters, our trustees and other Full Fact volunteers. Full details of our funding are available on our website.

We would welcome any thoughts or comments to our Chief Executive Chris Morris, at chris.morris@fullfact.org

Nuffield Foundation

[1] Full Fact considers bad information to be information that causes harm or promotes misunderstanding. It is often described as misinformation and/or disinformation, and we set out our definition of those terms later in this report.

Summary

Our information environment is changing with extraordinary speed. The emergence of popular generative AI tools has created new challenges and opportunities for anyone fighting against misinformation and disinformation. As an election approaches, politicians need to restore public trust in our political system, and pledge to use generative AI responsibly. The government needs to strike the right balance between protecting against harms online and promoting freedom of expression.

Full Fact campaigns to ensure citizens have access to reliable information. It allows them to make informed choices about the things that matter to them, from voting in an election to protecting their personal health. We scrutinise all sides in any political debate impartially, and try to hold everyone in public office to high standards. Last year we published 776 fact checks and secured 130 corrections from newspapers, broadcasters, MPs and the Prime Minister.[2]

2023 was also the year in which generative AI began redefining the way information is created, produced and consumed. Change is happening very quickly and in an era when more and more people are beginning to question what they should believe, it is essential for politicians to campaign for honesty and transparency, and to protect freedom of expression. It is the only way to rebuild public trust.

Generative AI is making it harder to address misinformation and disinformation in effective ways

The information revolution is gathering pace, and widely available generative AI tools are forcing us to re-examine the rules which govern our information environment, from media literacy to legislation and regulation. AI can be an enormous force for good, and Full Fact has used it to build automated fact checking tools for fact checkers and journalists in more than 20 countries, helping them separate checkable fact from opinion in the torrent of information online.

But generative AI is also a serious threat, making it cheap, easy and quick to spread misinformation and disinformation, and creating content so plausible that it is impossible to judge quickly whether something is real or not. In advance of a UK general election, politicians should pre-empt public disquiet about the legitimacy of an election influenced by AI, and promise publicly to use generative AI responsibly in all their campaigning and other political activity. Not to do so would further degrade already historically-low trust in UK politics and institutions.

[2] Full Fact, ‘Full Fact in 2023’, December 2023, https://fullfact.org/blog/2023/dec/full-fact-in-2023/

Governments, in the UK and elsewhere, also need to do more to ensure that elected representatives are involved in setting the rules of this new information environment. Where necessary, this should include legislation to help protect people from harm. The Online Safety Act should be updated fundamentally or replaced with new legislation which addresses the challenges brought about by AI in an effective way.

Responsibility also lies with online platforms and search engines, which have amassed enormous power. They must become far more transparent about the data they collect, and work together to develop international technical standards that benefit citizens around the world.

Striking the right balance

Our society thrives when it promotes robust public debate, based on accurate facts. But there can be tension between protecting freedom of expression and dealing with potential online harms. Full Fact is robustly pro-free speech, but we believe debate and disagreement needs to be based on a body of shared facts (while acknowledging that there will always be grey areas), and on evidence that stands up to statistical scrutiny.

We are convinced that it is possible to challenge misinformation or misleading statements without restricting freedom of expression. That means content neutral solutions like transparent labelling or the promotion of trustworthy information are preferable to removing content by default.

Carefully crafted legislation and regulation can also play an important role in protecting people online. But—as we argue in this report—this should be done transparently, to ensure that online regulation and moderation is underpinned by widespread public trust.

Improving honesty and accuracy in public life

In the run up to the next UK election, politicians from all parties have an important role to play in restoring trust more generally. They should always seek to promote honesty and accuracy when speaking publicly, and in their behaviour.

The next parliament could be a turning point for higher standards, and increased public confidence that the fight against misinformation and disinformation starts from the top. There is a window to rebuild public trust, if we start now. But people will not forgive leaders who simply do not try.

Our recommendations

1. The next government should amend the Online Safety Act—or bring in new legislation—to better address harmful misinformation and disinformation, especially relating to health or when generated by AI, and media literacy.

2. The next government should build on existing regulatory principles to tackle AI-generated misinformation and disinformation.

3. The government should enable Ofcom to have regulatory oversight of online platform and search engine policies on generative AI and content.

4. Online platforms and search engines should voluntarily commit to establishing and improving their policies on AI-generated misinformation and disinformation before the end of the current parliamentary session, regardless of whether the UK government compels them to do so.

5. Technology companies should participate in international standards for indirect disclosure techniques and be transparent about the accuracy and reliability of detection tools used to moderate content and enforce policies.

6. The next government must ensure that researchers and fact checkers have timely access to data from online platforms and search engines about misinformation and disinformation on their platforms, and the impact of fact checks.

7. Online platforms and search engines should provide long-term funding for fact checking organisations, tools they need, and their networks.

8. The government must increase resources for media literacy now and to meet future demand.

9. Ofcom should work with online platforms and search engines to ensure that media literacy interventions are responding to the needs of UK citizens, and seen by as many people as possible.

10. The government should set out how it will work transparently with online platforms and search engines to challenge misinformation and disinformation during the next general election, including in the event of an information incident.

11. Political parties should commit publicly to transparent and responsible use of AI during elections.

12. The Procedure Committee should finish implementing agreed changes to Parliament's corrections system without further delay and reform the standards mechanisms for the next Parliament so that MPs who do not uphold the principle of honesty are held to account.

13. Ministers and government departments must provide evidence for what they say, and use public data in line with the Code of Practice for Statistics. This must be embedded in the Ministerial Code, and Parliament must hold Ministers to account when they fail to live up to these standards.

14. All political parties must commit to honest campaigning during the next election.

15. The next government should legislate to end deceptive campaign practices; introduce independent regulation of political advertising; and put the Ministerial Code on a statutory footing.

Full Fact's work is only possible thanks to the support of thousands of individuals across the country. **For updates and opportunities to take action against bad information, join us:** fullfact.org/signup

Part 1: Generative AI and the information environment

How to mitigate the new risks of misinformation and disinformation generated by AI

New technology can help to spread accurate information effectively. But as AI tools become widely adopted by both members of the public and political campaigners, new risks have emerged regarding the creation and dissemination of misinformation and disinformation.

The term generative AI (also called synthetic media) is used frequently in this report, and refers to machine learning models that can create new content, whether that is audio, text or video. Generative AI models are trained on large datasets so that they can predict the most likely response to prompts or questions based on the patterns in that data.[3]

The first part of this report explores what the government, regulators, technology companies and civil society need to do to protect our information environment. We analyse where the Online Safety Act has left individuals and our democracy vulnerable, and assess the government's initial approach to AI regulation.

We also argue for greater coherence of online platform and search engine policies on AI, and set out how these companies, publishers, fact checkers and others can continue to collaborate for the public benefit on building tools and common technical systems to address bad information at scale.

The regulator with responsibility for media literacy, Ofcom, is under-resourced to deliver and oversee provision of media literacy at the scale needed. We outline how evaluation, funding and research can support citizens to navigate the new information environment.

Finally, Part 1 of this report focuses on the upcoming general election. We lay out how the UK needs transparency and better planning for information incidents to protect future elections, and urge party leaders to give the public reasons to trust what they are doing with generative AI to win our votes.

[3] Ofcom, Future Technology and Media Literacy: Understanding Generative AI, February 2024, https://www.ofcom.org.uk/__data/assets/pdf_file/0033/278349/future-tech-media-literacy-understanding-genAI.pdf

Chapter 1: The Online Safety Act does not protect UK citizens from the harmful effects of misinformation and disinformation

The government must address the gaps in the UK online safety regime, particularly regarding harmful health information

Recommendation: The next government should amend the Online Safety Act—or bring in new legislation—to better address harmful misinformation and disinformation, especially relating to health or when generated by AI, and enhance media literacy.

The Online Safety Act became law in October 2023 and contains measures intended to improve online safety in the UK. This includes duties on internet platforms about having systems in place to manage harmful content on their sites, including illegal content.[4] The government claimed that the Act would fulfil a manifesto commitment to make the UK the "safest place in the world to be online while defending free expression",[5] but there are fundamental gaps in its provision. The Act is not fit for purpose.

Despite promises that the regulation would apply to "disinformation and misinformation that could cause harm to individuals, such as anti-vaccination content,"[6] there are only two explicit areas of reference to misinformation in the final Act.[7] One is that a committee should be set up to advise the regulator, Ofcom, on policy towards misinformation and disinformation, and how providers of regulated services should deal with it. The other is that Ofcom's existing media literacy duties should expand to cover public awareness of misinformation and disinformation, and the "nature and impact of harmful content". This is not good enough, given the scale of the challenge we face.

[4] UK Government, 'Online Safety Act: new criminal offences circular', January 2024, https://www.gov.uk/government/publications/online-safety-act-new-criminal-offences-circular

[5] UK Government, *Online Safety Bill: supporting documents*, 17 March 2022, https://www.gov.uk/government/publications/online-safety-bill-supporting-documents#what-the-online-safety-bill-does.

[6] UK Government, 'Online Harms White Paper: Full government response to the consultation', 15 December 2020, https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response.

[7] Online Safety Act 2023, ch. 50. https://www.legislation.gov.uk/ukpga/2023/50.

We know that bad information ruins lives. The Covid-19 pandemic in particular highlighted the very real harms that can result from misinformation and disinformation online, and how in times of crisis, information vacuums can fuel the spread of harmful misleading information. In that context, the government's decision to abandon its commitment to address non-criminal content that is harmful to adults is disappointing. There are understandable and justified concerns that tackling online misinformation will come at the expense of freedom of speech.[8] But in our 2023 report on health misinformation,[9] we argued that it is possible to balance the two. There are content neutral methods available to regulators to reduce the harm from misinformation, which means that removing content should rarely be necessary. These include promoting good information, such as the Covid-19 information centres on Facebook; having initiatives which introduce friction, such as read-before-you-share prompts introduced by X (formerly Twitter); and highlighting independent fact checking. This principle of finding the right balance should be central to any new legislation or amendments, and should be front of mind for Ofcom.

The terms misinformation and disinformation are used a lot in this report so it is worth defining briefly what we mean by them. Misinformation is information that is false or misleading and could cause harm, but has not been created with the intention of doing so. Disinformation is information that is "false and is deliberately created to harm a person, social group, organisation or country[10]." The Online Safety Act should have been a pivotal moment in the debate about how we tackle the harms they cause. But the Act does not address health misinformation encountered by adults, fails to set out how to protect citizens from home-grown electoral disinformation, and does nothing to counter the risks of fast spreading misinformation and disinformation that occurs during information incidents, such as terror attacks.

The Act was also an important opportunity to create a safe and stable starting point as we move into a landscape now further complicated by the new challenges of AI-generated misinformation and disinformation. In Chapter 2 we will outline how the Online Safety Act does not address the particular harms brought about by misleading information in the form of generative AI, and how changes to the Act are now even more urgent.

[8] UK Government, 'New protections for children and free speech added to internet laws', 28 November 2022, https://www.gov.uk/government/news/new-protections-for-children-and-free-speech-added-to-internet-laws.

[9] Full Fact, 'Online health misinformation in the UK', April 2023, https://fullfact.org/media/uploads/online_health_misinformation_in_the_uk_full_fact.pdf.

[10] For this and other useful definitions see: C. Wardle and H. Derakhshan, 'Information Disorder: Toward an interdisciplinary framework for research and policy making', Council of Europe, 27 September 2017, https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c.

In this chapter, we discuss the specific shortcomings of how the Act treats misinformation and disinformation, with particular reference to content about health. We set out recommendations for what this and future governments need to do to improve it, and we make recommendations for Ofcom on how it can best work in the existing framework.

Case study: Health misinformation

Health misinformation is false or misleading content that relates to physical or mental health conditions or symptoms, or medical treatments or interventions. This can take the form of medical misinformation or, in some contexts, involve misinformation linked to health statistics.

Health misinformation can harm people's physical and mental health and delay the provision of care.

In July 2022, Full Fact secured a commitment from the government to include explicit protections against health misinformation for adults in the Online Safety Bill. This written statement in the House of Commons by the then Secretary of State for Digital, Culture, Media and Sport Nadine Dorries included "Harmful health content that is demonstrably false, such as urging people to drink bleach to cure cancer" as well as "some health and vaccine misinformation and disinformation".[11]

This would have been a significant win. We live in an era of burgeoning health misinformation across topics as varied as fertility, heart diseases, cancer and vaccinations[12] and there is mounting evidence of the impact of health misinformation on individuals.[13]

[11] 'Priority content (Category 1 services need to address in their terms and conditions): Harmful health content that is demonstrably false, such as urging people to drink bleach to cure cancer. It also includes some health and vaccine misinformation and disinformation, but is not intended to capture genuine debate.' UK Parliament, *Online Safety Update* (written ministerial statement UIN HCWS194), 7 July 2022, https://questions-statements.parliament.uk/written-statements/detail/2022-07-07/hcws194.

[12] Full Fact, 'Online health misinformation in the UK', April 2023, https://fullfact.org/media/uploads/online_health_misinformation_in_the_uk_full_fact.pdf.

[13] The Council of Canadian Academies, 'Fault Lines', 2023, https://www.cca-reports.ca/wp-content/uploads/2023/02/Report-Fault-Lines-digital.pdf.

In 2022, Full Fact began seeing claims on social media which falsely linked the child flu vaccine to Strep A.[14] [15] In that same year the uptake of the flu vaccine among 2 and 3 year olds dropped considerably when compared with the last two years.[16]

Full Fact regularly sees misinformation about cancer risks, treatments and cures online, including claims about alternative therapies.[17] [18] [19] [20] Cancer Research UK has said that individuals seeking alternative therapies based on misinformation might postpone or decline evidence-based conventional treatments, which might otherwise prolong or even save a patient's life.[21]

Ultimately, the government went back on its promise to include provisions for health misinformation and disinformation in the Act, despite being warned by Full Fact about how health misinformation causes harm, including increasing vaccine hesitancy and increasing stigma and prejudice. In May 2023 we wrote to the secretary of state, in a letter[22] co-signed by a number of health charities and prominent medical professionals, warning that the language and ambition in the Online Safety Bill needed to be strengthened. But the UK public has been left with very limited protection from the legislation which finally emerged from a lengthy process.

One of the few mentions of health misinformation in the final Online Safety Act can be found in the list of priority content that is harmful to children, for example content that encourages the ingestion, inhalation or exposure to harmful substances. Unfortunately, the same is not true for the priority content for protecting adults online.[23]

[14] Full Fact, 'Strep A deaths are not dangerous new strain caused by flu vaccines', 4 January 2023, https://fullfact.org/health/strep-A-historic-deaths/.

[15] Full Fact, 'Study didn't link children's flu vaccine to strep A infections', 21 December 2022, https://fullfact.org/health/strep-a-nasal-flu-vaccine-study/.

[16] UK Health Security Agency, 'Concern over low rate of 2 to 3 year olds getting the flu vaccine', 30 November 2022, https://www.gov.uk/government/news/concern-over-low-rate-of-2-to-3-year-olds-getting-the-flu-vaccine.

[17] Full Fact, 'Tumours are not "there to save your life"', 28 July 2022, https://fullfact.org/health/cancer-tumour-causes/.

[18] Full Fact, 'Facebook post claiming lemons treat cancer better than chemotherapy is false', 15 December 2022, https://fullfact.org/health/lemons-and-cancer/.

[19] Full Fact, 'No solid proof cannabis oil can 'cure' cancer', 9 August 2022, https://fullfact.org/health/cannabis-oil-cure-cancer/.

[20] Full Fact, 'Rubbing hydrogen peroxide over your body every day does not treat cancer', 27 January 2022, https://fullfact.org/health/hydrogen-peroxide-cancer-treatment/.

[21] Cancer Research UK, 'Alternative therapies: what's the harm?', 27 April 2015, https://news.cancerresearchuk.org/2015/04/27/alternative-therapies-whats-the-harm/.

[22] Letter from Full Fact to the Secretary of State for Science, Innovation and Technology, 10 May 2023, https://fullfact.org/media/uploads/harmful_health_misinformation_and_the_online_safety_bill_%E2%80%94_letter_to_secretary_of_state_%E2%80%94_10_may_2023.pdf.

[23] Online Safety Act 2023, c.50, https://www.legislation.gov.uk/ukpga/2023/50/section/62/enacted.

The only stipulation in law is that online platforms and search engines are required to enforce their terms of service consistently, which only has an effect if they have policies in place prohibiting health misinformation. This is not a given, and is vulnerable to sudden change, as the story of Twitter/X shows.[24]

Ensuring good health information online is not just about taking down the very worst material. It's also about implementing policies that enable users to make informed choices and that create signposts to good information on health.

We urge whoever leads the next government to strengthen regulation of online platforms and search engines, especially on health misinformation, as a matter of urgency. This includes requiring them to undertake adult risk assessments, to have clear policies on how they will tackle health misinformation, and to play a productive role in media literacy campaigns.

The next government and parliament must create better law and enable effective regulation on harmful misinformation

Anyone who has experienced the harm done by misinformation, either personally or at close quarters, will be disappointed by this government's failed promise to make the UK a safe place to be online—especially now it is apparent that this failure has put the UK in a worse position to contain new harms brought about by AI generated disinformation, then shared unintentionally as misinformation.

However, it is clear that a different future is possible, as the regimes introduced by other governments (set out in Chapters 2 and 3) tentatively indicate. Whoever forms the next government must face up to the challenge of improving our information environment and future-proofing regulation so that the risks of existing and new technologies do not go unchecked.

In its Public Communication Scan of the United Kingdom published in December 2023, the OECD highlighted "a noteworthy gap in the legislative and policy landscape [...] on mis- and disinformation in the context of elections", and recommended that the UK Government develop a comprehensive strategy on its present and future policy agenda to address this.[25]

Regardless of party makeup, the next government will need to revisit the Online Safety Act to ensure that Ofcom is able to tackle the harms of mis-information and disinformation during the next decade.

[24] New York Times, 'The consequences of Elon Musk's ownership of X', 27 October 2023, https://www.nytimes.com/interactive/2023/10/27/technology/twitter-x-elon-musk-anniversary.html.

[25] OECD, 'Public Communication Scan of the United Kingdom', 16 December 2023, https://www.oecd-ilibrary.org/sites/bc4a57b3-en/1/3/3/index.html?itemId=/content/publication/bc4a57b3-en&_csp_=0aa641c3d4fda7ac26451f2c0133d8cf&itemIGO=oecd&itemContentType=book.

Labour has said that, if it forms the next government, it will legislate as soon as possible. It says it intends to increase Ofcom's power to ensure that companies are held to account beyond enforcing their own terms and conditions. This would be an essential step—but only a first step—towards filling the gaps in the UK online safety regime.

With the Online Safety Act passed and its implementation in progress, the Conservatives are now turning their attention to AI. However, as we set out in Chapter Two, misinformation is not being prioritised in this context either. For example, the UK's AI Safety Summit in November 2023 focused on the medium to long-term future risks or loss of control presented by "frontier AI", such as biological or cyber-attacks, development of dangerous technologies, or critical system interference, rather than the generative AI-boosted risks of misinformation that are already here.[26]

Ofcom must make the best of a bad hand

In the meantime, Ofcom's power to regulate misinformation and disinformation is limited to the small remit set out in the Online Safety Act, and will continue to be limited without legislative changes.

Ofcom's research functions give it the ability to gain a deep understanding of the scale, scope and harm of misinformation and disinformation.[27] Based on its track record of quickly gaining an understanding of this area[28] [29] [30], we expect Ofcom to be willing and able to make confident public recommendations about whether or not online platforms and search engines are effectively and proportionately addressing the most harmful misleading content.

Ofcom should ensure that its forthcoming Advisory Committee on Disinformation and Misinformation draws on the full range of expertise from civil society and technologists—in the UK and beyond—in order to monitor and prioritise emerging challenges.

This modest recognition of the need to address misinformation and disinformation provides a starting point, and there is still time to increase Ofcom's powers and corresponding resources so that the regulator can much more effectively limit harm to individuals, groups and society.

[26] UK Government, 'AI Safety Summit: introduction', 31 October 2023, https://www.gov.uk/government/publications/ai-safety-summit-introduction/ai-safety-summit-introduction-html.

[27] Online Safety Act 2023, ch. 50, Chapter 7, https://www.legislation.gov.uk/ukpga/2023/50/part/3/chapter/7/enacted.

[28] Ofcom, 'Misinformation: A Qualitative Exploration', June 2021, https://www.ofcom.org.uk/__data/assets/pdf_file/0010/220402/misinformation-qual-report.pdf.

[29] Professor Lee Edwards et al., 'Rapid Evidence Assessment on Online Misinformation and Media Literacy', June 2021, https://www.ofcom.org.uk/__data/assets/pdf_file/0011/220403/rea-online-misinformation.pdf.

[30] Ipsos, 'Understanding experiences of minority beliefs on online communication platforms', September 2023, https://www.ofcom.org.uk/__data/assets/pdf_file/0019/268102/understanding-experiences-minority-beliefs.pdf.

We note that Ofcom has said the Advisory Committee will be set up by the end of 2024,[31] but we know of no reason why it could not be established sooner—and it should be.

Action for the government

- Develop a comprehensive strategy for UK action against misinformation and disinformation, including mechanisms for evaluation, scrutiny and accountability.
- Strengthen existing or introduce new legislation and regulation on harmful misinformation and disinformation, including addressing serious gaps on harmful health misinformation. This includes requiring online platforms and search engines to undertake adult risk assessments and media literacy programmes, and to have clear policies on how they will tackle health misinformation.

Action for Ofcom

- Use full research powers to gain deep understanding of all issues involving harmful online misinformation and disinformation, and make evidence-based recommendations on how the regulatory framework should be improved.
- By mid-2024, in time for an autumn election, set up the Advisory Committee on Disinformation and Misinformation which draws on expertise across the field in order to effectively monitor and prioritise emerging and existing harms.

[31] Joint Committee on the National Security Strategy, Oral evidence: Defending democracy, 18 March 2024, https://committees.parliament.uk/oralevidence/14513/html/.

Chapter 2: The UK approach to AI regulation must address harmful misinformation and disinformation effectively

Government action on AI must be coherent in order to build a good information environment

Recommendation: The next government should build on existing regulatory principles to tackle AI-generated misinformation and disinformation.

AI is accelerating the creation and distribution of misleading content

Artificial Intelligence, and in particular generative AI, can be an enormous force for good. It can help promote high-quality information, and encourage freedom of expression. However, Full Fact also has significant concerns about how generative AI tools will be employed in the creation of misinformation and structured disinformation campaigns.

It is becoming faster, easier, and cheaper to create highly sophisticated manipulated or synthetic (i.e. it looks real but is artificial) text, images, video and audio. At the same time, it is becoming more challenging to determine the authenticity of any material. This poses a range of new challenges.

AI makes success easier for those already set on leading disinformation campaigns, by putting new tools into their hands and allowing far more realistic content to be generated at much higher speed.[32] In the last few years, we have rapidly moved from a world where creating deepfakes needed huge amounts of computational power and skill,[33] to one where they are now a consumer product that can be created easily on a smartphone app.

[32] NewsGuard, *Tracking AI-enabled Misinformation: 766* 'Unreliable AI-Generated News' Websites (and Counting), Plus the Top False Narratives Generated by Artificial Intelligence Tools* (website), https://www.newsguardtech.com/special-reports/ai-tracking-center (*accessed 21 March 2024, the number of websites is likely to increase prior to publication of this report).

[33] Ian Goodfellow et al., 'Generative Adversarial Nets', Cornell University, 10 June 2014. https://arxiv.org/abs/1406.2661.

The sheer plausibility of the content being produced is also increasing the chances of misleading and false information being unintentionally spread. Even industry experts are unable to decide definitively if some content has been produced by humans or machines.[34] This means that reliable automated detection[35] is a long way from being something we can depend on. The recent announcement of the OpenAI Sora project[36] for the automated production of hyper-realistic video content suggested that this is a multifaceted challenge, with significant advances in the quality of AI-produced content happening in the space of a few months.

There are more misinformation-related risks that stem from the technology itself—for example, generative AI is known to create highly plausible but inaccurate content without users in the chain of dissemination being aware. These are known as 'hallucinations'.[37] In written text, this type of content can include the creation of false citations[38] that offer the aura of plausibility around even outlandish content, and potentially enable it to spread further.[39]

Freedom House reported in 2023 that "AI-based tools that can generate images, text, or audio were utilised in at least 16 countries to distort information on political or social issues".[40] That number is set to rise dramatically. The emergence of significant volumes of fake or low-quality content created in order to sow confusion could be intended less to promote belief in an individual claim, and more to reduce trust in information generally.

[34] Malay Mail, 'Is the political aide viral sex video confession real or a Deepfake?', 12 June 2019, https://www.malaymail.com/news/malaysia/2019/06/12/is-the-political-aide-viral-sex-video-confession-real-or-a-deepfake/1761422.

[35] Full Fact, 'No evidence clip of Sadiq Khan supposedly calling for "Remembrance weekend" to be postponed is genuine', 10 November 2023, https://fullfact.org/news/khan-audio-palestinian-remembrance/.

[36] Open AI, *Sora* (website), https://openai.com/sora (accessed 21 March 2024).

[37] Google Cloud, *What are AI hallucinations?* (website), https://cloud.google.com/discover/what-are-ai-hallucinations (accessed 21 March 2024).

[38] When asked to fact-check the claim "1 cup of dandelion greens = 535% of your daily recommended vitamin K and 112% of vitamin A", ChatGPT gave a correct verdict but fabricated a fake USDA study as supporting evidence. From: M. Schlichtkrull, Z. Guo, 'AVERITEC: A Dataset for Real-world Claim Verification with Evidence from the Web', University of Cambridge, 8 November 2023, https://arxiv.org/abs/2305.13117.

[39] Full Fact, 'Google snippets falsely claimed eating glass has health benefits', 15 November 2023, https://fullfact.org/health/google-snippet-eating-glass/.

[40] Shahbaz, Funk, and Vesteinsson, 'The Repressive Power of Artificial Intelligence,' in Shahbaz, Funk, et al. eds., 'Freedom on the Net 2023', Freedom House, 2023, https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence.

Generative AI also creates an environment in which real documentation of speech, actions and events can be easier to deny. It creates far more opportunities for genuine photojournalism images or audio, for example, to be dismissed as fake.[41] This 'liar's dividend'[42] could increase as more and more people become familiar with a growing number of tools that can make people appear to say or do things which in reality they have not done, further eroding trust in politics, institutions and other people online.

Not knowing if content is being produced for mischief or to influence an election will make it much harder for fact checkers, journalists and other organisations to monitor likely sources of disinformation, and to direct already limited resources in the most effective way.

Chapter 5 of this report discusses the pressing needs for higher quality technology, and for more information to be shared with fact checkers, to help address this. Chapters 3 and 4 outline the needs for online platforms, search engines and media organisations to invest more time into understanding how to best explain these concepts to the widest possible audience.

The UK Government has struggled to engage with these issues at the pace they have evolved

The Online Safety Bill was still going through Parliament as the first generative AI tools emerged and new users signed up in vast numbers.[43] Rapid adoption of these new technologies created a potential risk that the government chose not to address, when it could have put mechanisms in the Bill that would have helped provide oversight, such as insisting that online platforms and search engines undertake adult risk assessments.

The government said the legislation was technology-agnostic, and that content generated by AI would be covered as would any features using AI.

[41] Rest of World, 'An Indian politician says scandalous audio clips are AI deepfakes. We had them tested', 5 July 2023, https://restofworld.org/2023/indian-politician-leaked-audio-ai-deepfake/.

[42] R. Chesney and D. Keats Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security', 07 California Law Review 1753 (2019), U of Texas Law, Public Law Research Paper No. 692, U of Maryland Legal Studies Research Paper No. 2018-21, July 14, 2018, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954.

[43] Reuters, 'ChatGPT sets record for fastest-growing user base - analyst note', 2 February 2023, https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/.

> The Online Safety Bill has been designed to be technology-neutral to future-proof it and to ensure that the legislation keeps pace with emerging technologies. It will apply to companies which enable users to share content online or to interact with each other, as well as search services. Content generated by artificial intelligence 'bots' is in scope of the Bill, where it interacts with user-generated content, such as on Twitter [now X]. Search services using AI-powered features will also be in scope of the search duties outlined in the Bill.
>
> **Lord Parkinson, minister responsible for the Online Safety Bill in the House of Lords, February 2023[44]**

Practically, there is some coverage, but the overall impact is limited. Content that is generated by AI is treated in the same way as any other content, which means it is only regulated if it is in scope of the Ofcom regulatory regime. As we outlined in Chapter One, most harmful misinformation is not.

AI regulation misses an opportunity to focus on harmful misinformation

Since the government published its AI White Paper in March 2023[45], there has been significant public debate about its proposed approach. Now the government has also published its White Paper consultation response, titled "A pro-innovation approach to AI regulation"[46]. In this, there is a clear short-term focus on innovation funding and a suggestion of limited legislation later on. The consultation response proposes "targeted binding requirements on the small number of organisations developing highly capable general-purpose AI systems, to ensure that they are accountable for making these technologies sufficiently safe". In this instance, general-purpose AI systems are defined vaguely: "Foundation models that can perform a wide variety of tasks and match or exceed the capabilities present in today's most advanced models. Generally, such models will span from novice through to expert capabilities with some even showing superhuman performance across a range of tasks." It is not clear if or how this is intended to cover misinformation and disinformation.

[44] UK Parliament, written answer, Lord Parkinson of Whitley Bay, 17 February 2023, HL5570. https://questions-statements.parliament.uk/written-questions/detail/2023-02-08/hl5570.

[45] UK Government, 'AI regulation: a pro-innovation approach', 29 March 2023, https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach.

[46] UK Government, 'A pro-innovation approach to AI regulation: government response', 6 February 2024, https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response.

For now, the government says: “We are going to take our time to get this right—we will legislate when we are confident that it is the right thing to do.”[47] But, as we argue above, the possible outlines of an effective framework for addressing harmful misinformation and disinformation in a world dominated by AI have been obscured by the government’s intention to set its sights at a higher level. This vital issue is not being treated as a priority.

The government’s 2023 White Paper describes a framework to govern how regulators—ranging from the MHRA and the Care Quality Commission to the Information Commissioner’s Office (ICO) and the National Data Guardian—use their existing powers. The paper proposes five cross-sectoral principles for existing regulators “to interpret and apply within their remits in order to drive safe, responsible AI innovation”.[48] The principles cover the right areas: Safety, security and robustness; Appropriate transparency and explainability; Fairness; Accountability and governance; and, Contestability and redress. But the government needs to give more guidance to regulators on what they mean in practice, and ultimately translate this into law so that these principles stick.

Meanwhile, the AI Safety Institute—“the first state-backed organisation focused on advanced AI safety for the public interest”—has begun research into large language models (LLMs).[49] [50] It has already found that safeguards against disseminating harmful information are inadequate. However, the Institute’s remit is still developing, and it is not clear whether misinformation and disinformation safety risks will be a priority.

Ofcom will be responsible for online platforms and search services but, as we have argued in Chapter One, only in the very limited ways which apply to misinformation and disinformation that are set out in the Online Safety Act.[51] Like other UK regulators, Ofcom is required to publish its strategic approach for AI regulation by 30 April 2024,[52] with a 12-month roadmap, as well as an assessment of the risks and challenges in its sector, and a plan to address them. Ofcom has a wider remit on misinformation and disinformation when it comes to media literacy, and it should use this remit to substantively engage with how those specific risks intersect with AI (Chapter 6).

[47] UK Government, ‘A pro-innovation approach to AI regulation: government response’, 6 February 2024, https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response.

[48] UK Government, ‘AI regulation: a pro-innovation approach’, 29 March 2023, https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach.

[49] UK Government, ‘Introducing the AI Safety Institute’, November 2023, https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute.

[50] UK Government, ‘AI Safety Institute approach to evaluations’, 9 February 2024, https://www.gov.uk/government/publications/ai-safety-institute-approach-to-evaluations/ai-safety-institute-approach-to-evaluations.

[51] Ofcom will also cover AI in other areas it is responsible for, such as broadcast.

[52] UK Government, ‘A pro-innovation approach to AI regulation: government response’, 6 February 2024, https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response.

In the European Union (EU), the European Digital Media Observatory[53] has been set up to act as an independent convening body to facilitate fact checkers, academics, media organisations and media literacy experts to share information and act in a more coordinated manner to address misinformation and disinformation. Given the proliferation of AI-produced content, a similar body in the UK could be a viable option to ensure that those engaged with regulation have access to the best information as quickly as possible.

It is hugely challenging to introduce regulatory frameworks in a fast moving environment, especially ones that can engage meaningfully with the complexity of the issues at hand. With online misinformation and AI, this leaves us with a familiar pattern in which secondary legislation[54] is likely to be introduced to enable a far more rapid response to the developments of technology. But that means there will be a far lower level of scrutiny by our elected representatives of important changes to the law, and such action should not be the norm. Any work to legislate content online carries the risk of damaging people's freedom of speech, and such work should always be undertaken in the most transparent and open way possible, to avoid this happening.

Other governments have already taken steps to regulate AI systems

The EU has taken the first steps to regulate AI systems in law. Its Digital Services Act (DSA) had already gone further on misinformation and disinformation than any UK legislation. Now, with the EU's Artificial Intelligence Act (AI Act) agreed early in 2024, a legal framework for the regulation of AI systems is being brought in across the EU.

The EU AI Act includes provisions on "AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful".[55] There are transparency obligations for providers and deployers of certain AI systems and models, including disclosure requirements. Any UK government is highly unlikely to want to regulate in exactly the same way as the EU, but the AI Act sets a standard to which future UK regulatory efforts will inevitably be compared. There is justifiable concern that misinformation and disinformation will continue to fall between the gaps in the existing regulatory landscape.

[53] European Commission, *European Digital Media Observatory (EDMO)* (website), .https://digital-strategy.ec.europa.eu/en/policies/european-digital-media-observatory (accessed 21 March 2024).

[54] From a speech by the Permanent Secretary to the Department for Digital, Culture, Media andSport, Sarah Healey CB: "The future challenges for digital policy making in HMG", delivered at King's College London, 16 November 2022, https://thestrandgroup.kcl.ac.uk/wp-content/uploads/Sarah-Healey-speech-for-publication-v2.pdf#page=6.

[55] European Union Artificial Intelligence Act 2024, Art. 3(44bl), https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html.

In the US, the Biden administration has secured voluntary commitments from seven large companies, including Google, Meta, Microsoft and OpenAI, “to help move toward safe, secure, and transparent development of AI technology”.[56] These commitments include technical mechanisms “to ensure that users know when content is AI generated, such as a watermarking system” to reduce deception.[57] Voluntary agreements are of course not the same as binding legislation, and it remains to be seen how effective this system will be in that context; and as argued elsewhere, the UK will ultimately need some form of regulation.

The UK is also part of bilateral and multilateral partnerships and intergovernmental processes on AI, including at the G7,[58] G20, Council of Europe,[59] OECD,[60] United Nations, the Global Partnership on Artificial Intelligence (GPAI)[61] and the bi-annual AI Safety Summits.[62] However, no amount of international collaboration will disguise the UK Government’s failure to introduce adequate regulation to address the way that platforms treat harmful misinformation and disinformation. If this cannot be dealt with in the existing regulatory framework in the UK, we may have reached a point at which more foundational changes are required in the remit of existing regulators; or potentially, the introduction of a new regulator with a specific remit should be considered.

[56] The White House, ‘Fact Sheet: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI‘, 21 July 2023, https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.

[57] The White House, ‘Fact Sheet: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI‘, 21 July 2023, https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.

[58] OECD, ‘G7 Hiroshima Process on Generative Artificial Intelligence (AI) - Towards a G7 Common Understanding on Generative AI’, 7 September 2023, https://www.oecd.org/publications/g7-hiroshima-process-on-generative-artificial-intelligence-ai-bf3c0c60-en.htm.

[59] Council of Europe, ‘Council of Europe and Artificial Intelligence’, Council of Europe Publishing, March 2023, https://www.coe.int/en/web/artificial-intelligence/home.

[60] OECD, ‘G7 Hiroshima Process on Generative Artificial Intelligence (AI) - Towards a G7 Common Understanding on Generative AI’, 7 September 2023, https://www.oecd.org/publications/g7-hiroshima-process-on-generative-artificial-intelligence-ai-bf3c0c60-en.htm.

[61] *Global Partnership on Artificial Intelligence* (website), https://gpai.ai/ (accessed 21 March 2024).

[62] The Republic of Korea has agreed to co-host a mini virtual summit on AI in the 6 months after this Summit, and France will then host the next in-person Summit 6 months after that. Source: *AI Safety Summit* (website), https://www.aisafetysummit.gov.uk/ (accessed 21 March 2024).

The UK needs legislation to tackle the harms of AI-generated misinformation and disinformation, and the regulators delivering this regime must be sufficiently resourced

The government can still drive forward its intended work on AI and widen its focus to include misinformation and disinformation measures. However, the UK will need its own legislation to cover these issues, and this needs to be part of a coherent overall strategy to protect the UK public from harm, in an online information space which continues to change at sometimes bewildering speed.

While still possible, it seems unlikely that significant legislation will be brought forward ahead of the upcoming general election. The next government has been left with the big decisions on AI, and cannot simply rely on existing cross-sector regulatory principles as a safety net for tackling harmful misinformation and disinformation. These principles must be brought into legislation and regulation formally, with more detail on what they mean.

Labour continues to push the message of moving from a voluntary to a statutory system, for example the idea of forcing companies to share testing data under a statutory code rather than the existing voluntary one.[63][64] It is possible that Labour is considering a legislative vehicle to achieve this, but it has yet to publish its AI strategy.[65] Labour has also announced that it would introduce a Regulatory Innovation Office,[66] with the power to steer regulators' priorities. Theoretically, this office could bring pressure to bear on Ofcom's targets for action on misinformation and disinformation, or at minimum it could increase transparency about what the regulator is doing and whether this is working.

The Conservatives have not completely ruled out legislative backing for regulation. In its response to the AI White Paper,[67] the government says that "the challenges posed by AI technologies will ultimately require legislative action in every country once understanding of risk has matured". The government also promises that it "will shortly launch a call for evidence on AI-related risks to trust in information" and related issues such as deepfakes. This must happen urgently, as the election looms ever closer. While the government

[63] The Guardian, 'Labour would force AI firms to share their technology's test data', 4 February 2024, http://www.theguardian.com/technology/2024/feb/04/labour-force-ai-firms-share-technology-test-data.

[64] UK Government (AI Safety Summit), 'Safety Testing: Chair's Statement of Session Outcomes', 2 November 2023, https://www.gov.uk/government/publications/ai-safety-summit-2023-chairs-statement-safety-testing-2-november/safety-testing-chairs-statement-of-session-outcomes-2-november-2023.

[65] Computer Weekly, 'Labour will use AI to grow the economy by 0.5%, says shadow tech secretary Peter Kyle', 12 March 2024, https://www.computerweekly.com/news/366573312/Labour-will-use-AI-to-grow-the-economy-by-05-says-shadow-tech-secretary-Peter-Kyle.

[66] The Labour Party, 'Labour will end regulatory backlogs to give the public access to life-saving treatments sooner', 28 October 2023, https://labour.org.uk/updates/press-releases/labour-will-end-regulatory-backlogs-to-give-the-public-access-to-life-saving-treatments-sooner/.

[67] UK Government, 'A pro-innovation approach to AI regulation: government response', 6 February 2024, https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response.

drags its feet and waits for its understanding of AI risks to mature, the Data Protection and Digital Information Bill already has three amendments on deepfakes tabled (at the time of writing), including one from Labour on the "offence of creating or sharing political deepfakes".[68]

Finally, Ofcom has moved fast to hire relevant experts, so that it can develop its new role as effectively as possible given the resources it has. If Ofcom were to take on the regulation of harmful AI-generated misinformation and disinformation, this would require further boosts to its staffing capacity and expertise, and consequently more resources would be needed. The £10 million announced to upskill regulators on AI leadership is not sufficient.[69] The government should be clear about what proportion of this the Electoral Commission, Ofcom and the ICO should expect to get.

This government and the next government need to make urgent decisions about where the use of AI should be more, or less, strictly controlled in order to protect free speech, while building trust in online information and providing safeguards for citizens. Those decisions should be made a priority in the next parliament.

Action for the government

- Build on existing regulatory principles to regulate in the AI space to reduce the harms done by AI-generated misinformation and disinformation, and amend/replace the Online Safety Act.
- Explore the creation of a new independent digital observatory to facilitate collaboration between fact checkers, researchers, media companies and media literacy experts to respond to emerging technologies which have an impact on the information environment, such as AI.

Action for Parliament

- The House of Commons Science, Innovation and Technology committee should hold an inquiry on AI generated misinformation and disinformation.

Action for Ofcom

- Use media literacy remit to substantively research and consult on how misinformation and disinformation risks intersect with those posed by AI.

[68] From a search for "deepfake" on the UK Parliament webpage for the Data Protection and Digital Information Bill (Session 2023-24), https://bills.parliament.uk/bills/3430/stages/18402/amendments?searchTerm=%22deepfake%22&Decision=All (accessed 21 March 2024).

[69] UK Government, 'UK signals step change for regulators to strengthen AI leadership', 6 February 2024, https://www.gov.uk/government/news/uk-signals-step-change-for-regulators-to-strengthen-ai-leadership.

Chapter 3: The policies of online platforms and search engines on generative AI content must address bad information effectively

Greater coherence of policies and penalties should be combined with proper oversight in the UK

Recommendations: The government should enable Ofcom to have regulatory oversight of online platform and search engine policies on generative AI and content. Online platforms and search engines should voluntarily commit to establishing and improving their policies on AI-generated misinformation and disinformation before the end of the current parliamentary session, regardless of whether the UK government compels them to do so.

An overview of online platform and search engine policies on AI content

The widespread uptake of generative AI tools has prompted changes in the policies of online platforms and search engines. In the United States, the White House has taken a voluntary approach to this via an Executive Order which asks companies to develop and deploy mechanisms that enable users to understand if audio or visual content is AI-generated.[70]

In the European Union, signatories of the EU Code of Practice on Disinformation,[71] including Google, Meta, Microsoft and TikTok, were asked to "establish or confirm their policies in place for countering prohibited manipulative practices for AI systems that generate or manipulate content, such as warning users, and proactively detect such content".[72]

[70] The White House, 'Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence', 30 October 2023, https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

[71] European Commission, 'Signatories of the 2022 Strengthened Code of Practice on Disinformation', 16 June 2022, https://digital-strategy.ec.europa.eu/en/library/signatories-2022-strengthened-code-practice-disinformation.

[72] European Commission, '2022 Strengthened Code of Practice on Disinformation', 16 June 2022, https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation.

The EU's Digital Services Act (DSA) identifies a range of organisations that could be seen to have a role in creating a safe online environment. This includes internet access providers, domain name registrars, hosting services, online marketplaces, app stores, search engines and social media platforms.[73] It particularly identifies 'VLOPs'—very large online platforms with a reach of more than 10% of the 450 million consumers in the EU—as posing particular risks in terms of dissemination of illegal content and societal harms.[74] Since the introduction of the Code of Practice and the DSA, the EU's Artificial Intelligence Act has also been agreed, but it is not yet in force.[75]

Reviewing the different companies' approaches to moderating use of AI and AI-generated content (set out in the table below), it is clear there is a huge variety. Even within companies there is a lack of consistency—a view that is shared by others who have undertaken similar analysis.[76] For example, Meta has a clear policy prohibiting certain AI-generated content and has published information about its work with Partnership on AI to develop common technical standards for identifying AI content—yet WhatsApp's policy simply warns users that "Some images generated by AIs might not be accurate or appropriate",[77] [78] and its Help Centre commentary on generative AI is limited to cautioning users about the accuracy of AI rather than stipulating what content is allowed.[79] Similarly, YouTube and Google Play have policies on AI, while Google Search stops short at prohibiting "representation of actions or events that verifiably didn't take place"—and does not explicitly mention AI-generated or synthetic content.[80]

The varied and incomplete responses to a voluntary system in the US, and to the relatively structured and powerful regulatory system in the EU, suggests that outsourcing decisions to technology companies is not going to work in the UK. The government needs to set out minimum requirements for the policies of online platforms and search engines in legislation to ensure that citizens have a consistently safe and informed experience

[73] European Commission, *The Digital Services Act - ensuring a safe and accountable online environment* (website), https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en (accessed 21 March 2024).

[74] European Commission, 'Digital Services Act: Commission welcomes political agreement on rules ensuring a safe and accountable online environment', 23 April 2022, https://ec.europa.eu/commission/presscorner/detail/en/IP_22_2545.

[75] European Parliament, 'EU AI Act: first regulation on artificial intelligence', 8 June 2023, https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

[76] R. Miguel, 'Platforms' policies on AI-manipulated and generated misinformation', EU DisinfoLab, 6 December 2023, https://www.disinfo.eu/publications/platforms-policies-on-ai-manipulated-and-generated-misinformation/.

[77] Meta, 'Labeling AI-Generated Images on Facebook, Instagram and Threads', 6 February 2024, https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/.

[78] Whatsapp, *How to generate an AI image in a chat* (website), https://faq.whatsapp.com/666225138813752/?cms_platform=web (accessed 21 March 2024).

[79] Whatsapp, *How to generate an AI image in a chat* (website), https://faq.whatsapp.com/666225138813752/?cms_platform=web (accessed 21 March 2024).

[80] Google, *Content policies for Google Search* (website), https://support.google.com/websearch/answer/10622781?hl=en#zippy=%2Cmanipulated-media (accessed 21 March 2024).

when using the internet. At a minimum, this means: describing what is meant by 'AI-generated content'; explaining whether certain AI-generated content is prohibited or must simply be disclosed as being AI-generated; saying how companies will use disclosure information; and explaining what will happen if content and users fail to meet standards.

Good company policies on their own are not enough. They need to be matched with the resources to implement them properly, including sufficient levels of content moderation and human review for effective enforcement.

The policies of online platforms and search engines on generative AI

The table below presents examples of how a selection of online services, as broadly defined in the DSA, are setting rules about the creation and publication of AI-generated content, and how they say they enforce their own rules. This table is not exhaustive, and we do not comment on how effective these moderation and enforcement policies are.

The policies of online platforms and search engines on generative AI

Company or platform	Policy summary	What does it say on labelling and contextualisation?	What penalties does the company apply to users who do not comply?
Meta (Instagram, Facebook and Threads)	Prohibits synthesised videos which are likely to mislead an average person to believe subject said words they did not say; prohibits AI videos that merge or combine content to create a video that appears authentic.[81]	Meta says it will introduce labels for content that it detects are AI-generated,[82] and says it will require users to disclose that video or audio is AI-generated via a labelling tool when sharing content; high risk or publicly important matters may be labelled more prominently.	Remove content and/or reduce distribution; unnamed potential penalties for users who fail to disclose that content is AI-generated, including advertisers.[83]

[81] Meta, *Community Standards - Manipulated media* (website), https://transparency.fb.com/en-gb/policies/community-standards/manipulated-media/ (accessed 21 March 2024).

[82] Meta, 'Labeling AI-Generated Images on Facebook, Instagram and Threads', 6 February 2024, https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/.

[83] Meta, 'Helping people understand when AI or digital methods are used in political or social issue ads', 8 November 2023, https://www.facebook.com/government-nonprofits/blog/political-ads-ai-disclosure-policy.

Company or platform	Policy summary	What does it say on labelling and contextualisation?	What penalties does the company apply to users who do not comply?
YouTube	Asks users to disclose that they have created synthetic content that is realistic, including by using AI tools. Content technically manipulated / doctored that may pose a serious risk of egregious harm is not allowed on the platform.	YouTube has promised to add a new label to the description panel indicating that some of the content was altered or synthetic. Some content about sensitive topics may be labelled more prominently,[84]	If people consistently choose not to disclose that content is synthetic, content may be removed and they may be removed from the platform's monetisation programme. The platform has a three strikes system. Continually breaking community guidelines may lead to account suspension.
Google Play	Places responsibility on developers for ensuring generative AI apps do not generate offensive content; requires that apps that generate content must contain in-app flagging features for users and that developers should use these to inform content filtering and moderation in their apps.		App rejection or removal and freezing in-app purchases; limited app visibility or limited regions; and restriction, suspension or termination of developer account for multiple violations (which affects users ability to see and use other apps from the same developer).
TikTok	Requires users to disclose synthetic media that depicts realistic scenes; prohibits synthetic media that contains the likeness of any real private figure or adult public figures when used in the context of political or commercial endorsements.[85]	Users can add labels including stickers or captions.[86]	Not disclosed within policy.
X (formerly Twitter)	Prohibits synthetic media that may deceive or confuse people and lead to harm.[87]	Content that is significantly or deceptively fabricated may be labelled.	Post deletion, reducing visibility or users' ability to engage. Accounts continually sharing or advancing may be locked or suspended.

[84] YouTube, 'Our approach to responsible AI innovation' 14 November 2023, https://blog.youtube/inside-youtube/our-approach-to-responsible-ai-innovation/.

[85] TikTok, *Community Guidelines - Synthetic and Manipulated Media* (website), https://www.tiktok.com/community-guidelines/en/integrity-authenticity/#3 (accessed 21 March 2024).

[86] TikTok, 'New labels for disclosing AI-generated content', 19 September 2023, https://newsroom.tiktok.com/en-us/new-labels-for-disclosing-ai-generated-content.

[87] X, *Synthetic and manipulated media policy* (website), https://help.twitter.com/en/rules-and-policies/manipulated-media (accessed 21 March 2024).

Company or platform	Policy summary	What does it say on labelling and contextualisation?	What penalties does the company apply to users who do not comply?
LinkedIn	Prohibits synthetic media that distorts real-life events and is likely to cause harm.[88]	Linkedin makes no reference to labelling within its policies.	Removal and disabling distribution beyond the author's network.

Improvements to be made

Create clear policies where they do not currently exist

Looking across the different policies of online platforms and search engines, it is clear there are gaps. Some companies simply have no policy about AI-generated content, while others are over-reliant on manipulated media policies, and there is a risk that these become overwhelmed or outdated. In any case, companies need to consider the particular risks to users posed by generative AI in their spaces and publish clear policies accordingly. Creators, developers and consumers deserve clarity about what companies find acceptable, and what to expect in terms of enforcement and penalties if rules are not followed—and companies need to be consistent and efficient about following their own rules once clear standards are in place. As highlighted in Chapter 4, part of this work is to ensure that online platforms agree on which common standards should be adopted around what's known as indirect disclosure.

Publish policies to protect elections from misleading AI-generated content

With elections being held in many countries around the world in 2024, it is essential that online platforms and search engines outline how they intend to mitigate the risks of misleading AI-generated content in the context of election campaigns, while paying due attention to different global contexts and the need to protect freedom of expression, particularly during election periods.

In February 2024, 20 leading online platforms and search engines, including Google, Meta, Microsoft, TikTok, and X pledged to work together to detect and counter harmful AI content, in an AI Election Accord[89] released at the Munich Security Conference. One of the commitments is to provide transparency to the public regarding how they address "deceptive AI election content", with an explicit mention of "publishing the policies that explain how we will address such content". TikTok, Meta, Google[90] and Microsoft[91] have

[88] LinkedIn, *False or misleading content* (website), https://www.linkedin.com/help/linkedin/answer/a1340752/ (accessed 21 March 2024).

[89] AI Elections Accord, *A Tech Accord to Combat Deceptive Use of AI in 2024 Elections* (website), https://www.aielectionsaccord.com/ (accessed 21 March 2024).

[90] 'Supporting the Elections for European Parliament in 2024', 9 February 2024, https://blog.google/around-the-globe/google-europe/supporting-elections-for-european-parliament-2024/.

[91] 'Meeting the moment: combating AI deepfakes in elections through today's new tech accord', 16 February 2024, https://blogs.microsoft.com/on-the-issues/2024/02/16/ai-deepfakes-elections-munich-tech-accord/.

addressed this commitment through published policies: others need to complete this important work by the time of the UK local and mayoral elections. [92] [93]

Health-specific policies for AI-generated content

Following the widespread recognition of the severe harm done by health misinformation during the pandemic, some companies now treat misleading health information as a special case. Meta highlights health misinformation as one of four specific types of misinformation eligible for removal within its general misinformation policy.[94] YouTube has a freestanding policy on medical misinformation,[95] and regularly publishes blogs on health information and medical misinformation.[96] Concerns have been raised by leading academics, as well as the World Health Organisation, about the interaction between generative AI and health misinformation,[97] for example the risks that this technology entrenches vaccine hesitancy by amplifying emotional drivers or encourages inadequate treatment recommendations.[98] [99]

It follows that companies with an existing commitment to promoting informed health choices should consider development of their policies on generative AI through the lens of health and clinical misinformation, and the need to enable users to make informed health choices.

Build on emerging consensus about labelling AI-generated content

More than ever before, it seems that there is an emerging consensus on the legitimacy of labelling and contextualising content, as a choice that is distinct from simply removing content. In its *Responsible Practices for Synthetic Media* framework,[100] Partnership on AI, which convenes academic, civil society, industry, and media organisations, has

[92] Meta, 'Helping people understand when AI or digital methods are used in political or social issue ads', 8 November 2023, https://www.facebook.com/government-nonprofits/blog/political-ads-ai-disclosure-policy.

[93] TikTok, *Community Guidelines - Synthetic and Manipulated Media* (website), https://www.tiktok.com/community-guidelines/en/integrity-authenticity/#3 (accessed 21 March 2024).

[94] Meta, *Community Standards - Misinformation* (website), https://transparency.fb.com/en-gb/policies/community-standards/misinformation/ (accessed 21 March 2024).

[95] YouTube, *YouTube Help - Medical misinformation policy* (website), https://support.google.com/youtube/answer/13813322?hl=en (accessed 21 March 2024).

[96] Source: a search for the term 'health' on the YouTube blog website, available at https://blog.youtube/search/?domain=youtube&tags=youtube-health&order=newest (accessed 21 March 2024).

[97] World Health Organisation, 'WHO calls for safe and ethical AI for health', 16 May 2023, https://www.who.int/news/item/16-05-2023-who-calls-for-safe-and-ethical-ai-for-health.

[98] H. Larson and L. Lin, 'Generative artificial intelligence can have a role in combating vaccine hesitancy', BMJ 2024;384:q69, https://www.bmj.com/content/384/bmj.q69.

[99] R. Hatem, B. Simmons, J.E. Thornton. 'A Call to Address AI "Hallucinations" and How Healthcare Professionals Can Mitigate Their Risks', Cureus, 2023, Sep 5;15(9):e44720, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10552880/.

[100] Partnership on AI, 'PAI's Responsible Practices for Synthetic Media - A Framework for Collective Action', 27 February 2023,https://partnershiponai.org/wp-content/uploads/2023/02/PAI_synthetic_media_framework.pdf.

recommended labelling as a useful intervention and a practice that should be undertaken by technology companies. Full Fact has advocated for labelling over many years, as a free speech response to misinformation. Now is the time to build on this consensus.

Gaps in the policies of online platforms and search engines mean UK government action is essential

The UK is at risk of falling behind when it comes to grappling with these new and emerging technologies. Companies are implementing a patchwork of mismatched actions to address the challenges and opportunities they see.

However, as shown in our analysis, we can't simply rely on the good work of those who have chosen to engage, when such significant gaps in policies continue to exist across multiple companies and topics. There is value in making sure a company lives up to the promises it makes to its users. But this does not give the proper democratic oversight on how information and content that is false will circulate to UK citizens, how freedom of expression will be upheld and whether the harms being done to our society and democracy—as well as the harms being done to individuals—are being addressed.

As long as no action is taken, the UK Government and legislators in Westminster are deferring to the commercial incentives of Silicon Valley, and political leadership from Brussels or Washington DC. While Ofcom will take an interest in these policies, there are currently limits to its regulatory room for manoeuvre.

Action for online platforms and search engines

- Publish a specific policy addressing the use of AI-generated content, including:
  - A definition of synthetic or AI-generated content.
  - Direction on whether certain AI-generated content is prohibited or must simply be disclosed as being AI-generated.
  - Explanation of how companies will use disclosure information.
  - Explanation of penalties applied to users and accounts which fail to meet standards.
  - The way in which the companies themselves will monitor how they are following their own standards.
- Ensure that there are sufficient resources for human involvement in moderation and enforcement to ensure effectiveness of policies.

Action for the government

- Establish UK regulatory oversight of the policies of online platforms and search engines on generative AI and content, in order to ensure that companies:
  - Define what synthetic or AI-generated content means.
  - Make clear whether certain AI-generated content is prohibited or must simply be disclosed as being AI-generated.
  - Explain how companies will use disclosure information.
  - Explain penalties applied to users and accounts which fail to meet standards.

- In the absence of related regulation before the general election, the government should seek voluntary commitments from companies to improve policies by the end of the current parliamentary session.

Chapter 4: Common technical standards can help build trust

Technology companies, publishers and governments must continue to collaborate on technical systems to address bad information at scale, while acknowledging current limitations

Recommendation: Technology companies should participate in international standards for indirect disclosure techniques, and be transparent about the accuracy and reliability of detection tools used to moderate content and enforce policies.

There are solutions in technology, but there are no silver bullets

The information environment is moving at a faster pace than at any point since the invention of the internet. Generative AI poses huge challenges in terms of our ability to detect and trace different types of media. Some solutions outlined in this chapter seem to be on the cusp of success, but these will require cooperation across the whole information system and investment to help tools reach maturity.

Those responsible for social media platforms, web browsers, app stores and messaging services can radically change the quality of information we consume, regardless of where we are consuming it. At the moment, though, the implementation of this vision is patchy.

As well as voluntary agreements secured by the Biden administration for technology companies to "commit to developing robust technical mechanisms",[101] some companies are also making independent moves to improve the information system. For example, Meta has said it will routinely introduce labelling of content to its platforms where it detects indicators saying that something is AI generated.[102]

Two solutions are gathering momentum successfully, and producing cooperation from many of the necessary actors: Synthetic Media Transparency Methods, and automatic detection tools. As part of their commitment to the AI Election Accord in February 2024,

[101] The White House, 'Ensuring Safe, Secure, and Trustworthy AI', July 2023, https://www.whitehouse.gov/wp-content/uploads/2023/07/Ensuring-Safe-Secure-and-Trustworthy-AI.pdf.

[102] Meta, 'Labeling AI-Generated Images on Facebook, Instagram and Threads', 6 February 2024, https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads.

many companies promised to limit the risks of AI during elections through prevention (meaning attaching "provenance signals to identify the origin of content where appropriate and technically feasible") and detection (meaning attempting to detect deceptive AI content or authentic content using these and other signals). There is little detail on how companies will put this into practice, but there is an expectation that steps will be taken throughout 2024.

Transparency methods are varied but more alignment is needed

Transparency technologies cover a wide spectrum and are intended to help systems to judge whether a piece of content is original, manipulated or synthetic and if it originally comes from a verified source.

These technologies do this by either storing information in the metadata of items of content like photos or videos at the point of capture; adding a 'signature' from the creator to verify that it is the original version of that bit of content; or by allowing creators to attach additional information during the editing process.

These signals can then be read by other platforms and the information can be displayed to users, setting out for example who created the content, when and how. As long as the service provider (e.g. social media platform) is using open standards, this should be a relatively simple process: widespread participation is the biggest challenge to its effectiveness.

Some of the components required for this system to work are starting to gain traction. One of the biggest umbrella organisations in this area is the Coalition for Content Provenance (known as C2PA), which brings together the Adobe-founded Content Authenticity Initiative and Project Origin (an initiative from Microsoft and the BBC to tackle online disinformation).[103] C2PA has made progress in developing and promoting uptake of technical standards specifically for provenance, with participants and steering committee members including some of the biggest and most powerful companies in the global information ecosystem, and including hardware manufacturers, online commerce stock image providers, software companies, news media and search and social companies.[104]

[103] Coalition for Content Provenance and Authenticity, *Overview* (website), https://c2pa.org (accessed 21 March 2024).

[104] Coalition for Content Provenance and Authenticity, *Membership* (website), https://c2pa.org/membership/ (accessed 24 March 2024).

Adobe should be given credit for its support and championing of the concept of content credentials, a form of content verification which allows people creating content to add metadata at the point of export or download. This helps consumers or users of that content to understand whether it has been adapted or manipulated.[105] The purpose of these credentials is to build trust. This approach is especially important as search engines and other types of internet service providers get to grips with AI spam.

Automated labelling of AI-generated content could lead to a potential fatigue in users if larger and larger volumes of content are created and it becomes the norm.

Partnership on AI, another leading convenor of AI stakeholders, has promoted open collaboration among online platforms such as Meta, Google and Microsoft to develop cross-industry markers that can be used to detect and act upon AI-generated content (for example by labelling or disclosing) regardless of where the content originated.[106] [107] [108] This type of collaboration is rarely seen or conducted so openly and should be applauded.

PAI has noted that when it comes to adopting transparency methods, "While many disparate efforts have emerged to help audiences navigate an increasingly synthetic information environment, largely by providing context about content, the community has not aligned on which combination of tactics to implement, when to share insights with audiences, and how they can evaluate their efficacy in supporting trustworthy content."[109]

The BBC and the Royal Society concluded in a recent report that "digital content provenance is an imperfect and limited—yet still critically important—solution to the challenge of AI-generated misinformation."[110] However, the current breadth of participation in these initiatives, and the public commitment, provides a good basis for the future of open provenance standards that can have huge global public benefit.

[105] Adobe, *Content Credentials* (website), https://helpx.adobe.com/creative-cloud/help/content-credentials.html (accessed 21 March 2024).

[106] Meta, 'Labeling AI-Generated Images on Facebook, Instagram and Threads', 6 February 2024, https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads.

[107] Partnership on AI, 'PAI's Responsible Practices for Synthetic Media - A Framework for Collective Action', 27 February 2023,https://partnershiponai.org/wp-content/uploads/2023/02/PAI_synthetic_media_framework.pdf.

[108] Partnership on AI, 'PAI Announces Google to Join Framework for Collective Action on Synthetic Media', 14 July 2023, https://partnershiponai.org/pai-announces-google-to-join-framework-for-collective-action-on-synthetic-media/.

[109] Partnership on AI, 'Building a Glossary for Synthetic Media Transparency Methods, Part 1: Indirect Disclosure', 19 December 2023, https://partnershiponai.org/glossary-for-synthetic-media-transparency-methods-part-1-indirect-disclosure/.

[110] The Royal Society, 'Generative AI, content provenance and a public service internet - summary note of a workshop held on 14 – 15 September 2022', July 2023,https://royalsociety.org/-/media/policy/projects/digital-content-provenance/Digital-content-provenance_workshop-note_.pdf.

Behind the scenes: content authenticity labels

It is possible for AI tools to mark content as being AI-generated automatically at the point of creation, both with markers that are visible to users (a traditional watermark for example) and invisible markers such as fingerprinting, signing, and invisible watermarks.

Markers are evolving and the recent Google Synth[111] project indicates they could become very hard to remove. Whichever technology becomes the most widely used, this should become codified into standards to ensure that contextual information can be displayed regardless of where content travels online.[112] Support for these techniques should be standard for all generative AI tools.

These approaches require the original creators of the content to signal both where content has been artificially created and, in the case of projects like the Content Authenticity Initiative[113], to be equally clear when content is in original, unedited form. Both methods are important to build trust in good content, and to show clearly where users need to exercise caution.

While the systems are continuing to evolve, content produced by authoritative organisations should be treated as a special case when declaring AI use and held to higher standards. Any form of synthetic content published by political parties, governments, national statistical offices, health organisations and similar should provide as many signals as practically possible to ensure users are left with no ambiguity about the role of technology in its production.

What users see: direct disclosure

Many of the technologies currently being developed are focused on indirect disclosure tools. Indirect disclosure is a technical signal for defining whether a piece of media was created by AI.[114] Direct disclosure is the other side of the coin: how these signals are then displayed to users alongside the content.[115]

[111] Google, *SynthID* (website), https://deepmind.google/technologies/synthid/ (accessed 21 March 2024).
[112] Meta, ‘Labeling AI-Generated Images on Facebook, Instagram and Threads’, 6 February 2024, https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads.
[113] *Content Authenticity Initiative* (website), https://contentauthenticity.org/ (accessed 21 March 2024).
[114] Partnership on AI, ‘Building a Glossary for Synthetic Media Transparency Methods, Part 1: Indirect Disclosure’, 19 December 2023, https://partnershiponai.org/glossary-for-synthetic-media-transparency-methods-part-1-indirect-disclosure/#Indirect_Disclosure.
[115] Partnership on AI, ‘Building a Glossary for Synthetic Media Transparency Methods, Part 1: Indirect Disclosure’, 19 December 2023, https://partnershiponai.org/glossary-for-synthetic-media-transparency-methods-part-1-indirect-disclosure/#Direct_Disclosure.

Of course, not all AI generated content will be misinformation or disinformation. However, it seems likely that those concerned with maintaining trust in information online will decide to directly disclose content that has been generated by AI. This may be a helpful step, but it will need to be backed up with investment in research that seeks to understand: how to present labels in a way that does not degrade trust in information that is not labelled but is nevertheless high quality or accurate; how to actively support users' understanding of the content they are consuming; and how to maintain users' privacy by not sharing identifiable information about individuals unnecessarily.

Ofcom already has duties to help users establish the reliability, accuracy and authenticity of content found on the services it regulates. It could build on this by investing in research to help inform online platforms and search engines how they can best support users' ability to judge content authenticity.

The pitfalls of detecting AI generated content

In Full Fact's day-to-day fact checking we have seen a small, but noticeable, rise in claims we check which are potentially being produced by AI—for example recent checks of an audio clip which purported to be the Labour leader Sir Keir Starmer verbally abusing his staff,[116] and another audio clip claiming to be the London mayor Sadiq Khan apparently saying "Remembrance weekend" should be postponed in favour of a pro-Palestinian march.[117] We found no evidence to suggest that either clip was genuine.

There are several types of apps available that allow people to create misleading content using AI. 'Lip-syncs' take a genuine video of someone and use AI to adjust their mouth to make it look like they're saying something else. These might be accompanied by audio that is AI-generated or created by an impersonator. 'Puppet master' deepfakes use AI to animate the entire head, like the young Tom Cruise parody on TikTok.[118]

It is very hard, even for professional audio experts, to tell whether audio is real or not—and if it's not real, how exactly it was faked. As AI technology advances, media literacy tips for spotting deepfakes will likely date fast. There are many tools that claim to be able to tell you, some with a specific percentage of confidence, whether an image or video was generated using AI. However, at the time of writing, Full Fact doesn't quote these tools in our articles because we find they don't work consistently.

In an interview with tech publication 404 media, Professor Hany Farid, a digital forensics expert and academic, pointed out the example of an apparently real image that surfaced

[116] Full Fact, 'No evidence that audio clip of Keir Starmer supposedly swearing at his staff is genuine', 11 October 2023, https://fullfact.org/news/keir-starmer-audio-swearing/.

[117] Full Fact, 'No evidence clip of Sadiq Khan supposedly calling for 'Remembrance weekend' to be postponed is genuine', 10 November 2023, https://fullfact.org/news/khan-audio-palestinian-remembrance/.

[118] TikTok, @*deeptomcruise* (website), https://www.tiktok.com/@deeptomcruise?lang=en, accessed 21 March 2024.

during the Israel-Gaza conflict purporting to show the burnt body of a baby. The image had been put into a free AI checker tool, which concluded—apparently wrongly—that the image was made with AI. The "black box" automated tools are "not very explainable", Professor Farid said.[119]

Mike Russell, founder of the audio production company Music Radio Creative and a certified audio professional with more than 25 years of experience, told Full Fact in an interview that tools alone can't confirm whether something is genuine or not. Some of the tests he ran on the alleged audio of Mr Starmer swearing suggested that it was actually real.[120]

A New York Times feature testing AI detectors also highlighted the opposite danger: of genuine images being falsely labelled as AI-generated.[121]

> Every time somebody builds a better generator, people build better discriminators, and then people use the better discriminator to build a better generator...The generators are designed to be able to fool a detector.[122]
>
> **Cynthia Rudin, a computer science and engineering professor at Duke University, and principal investigator at the Interpretable Machine Learning Lab**

Modern AI models are trained on over a billion data points and this volume helps them output more and more plausible content. Detecting the use of AI in content creation is therefore an incredibly hard technical challenge and even the largest and most well resourced organisations in the world struggle to resolve it. For example, Meta has begun to deploy a deepfake detection model which it attempts to keep up to date in real time by generating similar deepfake examples to the ones it has already detected. However, Meta has not revealed information about the success and accuracy of this system beyond admitting "there's much more work to do", and that the problem calls for "long-term investments and a coordinated effort from researchers, engineers, policy experts, and others across our company."[123]

[119] 404 Media, 'AI Images Detectors Are Being Used to Discredit the Real Horrors of War', 14 October 2023, https://www.404media.co/email/cc4c9f18-f02a-4ff0-ba93-0d1e8dd81ed6/.

[120] Full Fact, 'How to spot deepfake videos and AI audio', 20 December 2023, https://fullfact.org/blog/2023/dec/how-to-spot-deepfakes/.

[121] New York Times, 'How Easy Is It to Fool A.I.-Detection Tools?', 28 June 2023, https://www.nytimes.com/interactive/2023/06/28/technology/ai-detection-midjourney-stable-diffusion-dalle.html.

[122] New York Times, 'How Easy Is It to Fool A.I.-Detection Tools?', 28 June 2023, https://www.nytimes.com/interactive/2023/06/28/technology/ai-detection-midjourney-stable-diffusion-dalle.html.

[123] Meta, 'Here's how we're using AI to help detect misinformation', 19 November 2020, https://ai.meta.com/blog/heres-how-were-using-ai-to-help-detect-misinformation/.

OpenAI removed its text detection service from the internet because of low quality results,[124] whereas Intel claims a 96% accuracy rate from its detection tools.[125] We are not yet in a position to state confidently that the results generated by these tools will consistently deliver the accuracy and standard needed to build trust in the long term, because they are still in the early stages of development and the industry is moving so rapidly. It is vital that when these kinds of results are explained to a user, they are able to understand immediately whether it is being authoritatively stated that content has been AI-generated, or whether it is being suggested that it may have been AI-generated.

Consensus is needed on which technologies to get behind

It is clear that automating solutions to this problem on a mass scale is not yet possible. Neither transparency methods nor automated detection are a silver bullet, but both are developing rapidly and deserve to be the focus of considerable funding and commitment from online platforms and search engines. However, we must expect to see continuing human involvement in content moderation and enforcement systems for the foreseeable future.

Until there is more consensus across the technology industry, civil society and others about which technologies to coalesce behind, we need to look at other solutions such as media literacy (Chapter 6), legislation (Chapters 1 and 2) and the policies of online platforms and search engines (Chapter 3).

Action for Ofcom

- Undertake research and convene experts to build recommendations for online platforms and search engines about the effective and transparent presentation of disclosure information to users.

Action for technology companies

- Consolidate existing work to a small number of well defined common standards for indirect disclosure techniques, and ensure adoption across all parts of content production and distribution.
- Publish information about the accuracy and reliability of detection tools used in moderation and enforcement systems.
- Continue to participate in and support cross sector initiatives that promote international standards of interoperability that have global public benefit.
- Provide long-term investment in detection tools, whether or not these are used within company systems, and give these to fact checkers.

[124] OpenAI, 'New AI classifier for indicating AI-written text', 31 January 2023, https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text.

[125] Intel, 'Intel Introduces Real-Time Deepfake Detector', 14 November 2022, https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html.

Chapter 5: Ensure fact checkers have the tools and data needed to fight harmful misinformation and disinformation

The mainstream adoption of generative AI means fact checkers need more support to verify the accuracy of content and claims

Recommendations: The next government must ensure that researchers and fact checkers have timely access to data from online platforms and search engines about misinformation and disinformation on their platforms, and the impact of fact checks. These companies should provide long-term funding for fact checking organisations, tools they need, and their networks.

Help fact checkers find harmful misinformation and disinformation, and check it as quickly as possible

Fact checkers are often first responders, fighting on the front line against misinformation and disinformation. Over the last 15 years, fact checkers in more than 71 countries have collectively produced over 200,000 checks,[126] each one attempting to make the information environment better. This can be a challenging world to operate in, with often small teams constantly trying to adapt to the scale and evolution of the internet. Fact checkers need the best tools, technology, data and support to help them with this important task.

The speed of development of new methods for distributing online material, increasingly aided by AI-produced synthetic content, means that this challenge is greater than ever before. It is vital that fact checkers are supported with access to innovative tools which can help them, where possible, detect whether content has been AI-generated. It is also vital that fact checkers have access to high quality and timely information from within online platforms and search engines to help them understand as quickly as possible where harmful information and narratives are forming, and to ensure fact checks can make the greatest difference.

[126] *Fact-Check Insights* (website), https://www.factcheckinsights.org/ (accessed 21 March 2024).

Fact checking organisations also need to be able to improve their effectiveness—and prove it to existing and prospective funders. This requires a greater understanding of the impact that fact checks have, for example who sees fact checks or labels based on fact check metadata, and how this affects people's choices to consume or share information. At the moment, online platforms and search engines choose to share very little of this kind of information with fact checkers, even when they are working within formal partnerships. For example, Meta has released information about the impact of its Third-Party Fact-Checking programme globally and at an EU level,[127] but many fact checkers are concerned about whether their work is making any difference at a national level.

In correspondence with Meta and at public events, fact checkers have asked to see a breakdown of the number of items of content which have had fact check labels applied, on a country by country basis. This would help fact checking organisations to judge whether they should be investing so much effort in working with online platforms and search engines, or whether they should use their resources in other ways. Fact checkers have also asked for more granular and country-level information about whether seeing a fact check label means people are less likely to reshare posts which contain false information, and whether fact-check labels result in a reduction in likes or follows.

This is especially important in a world of generative AI and large language models (LLMs). Information produced in fact checks can be used as training data to enhance other products, and the results of fact checks can be shown to users in a moderated or personalised experience through interfaces like chat assistance, without fact checkers being aware that this is happening. This makes it even harder to track the reach of the original content.

Finally, fact checking organisations need support through sustained funding. This needs to come in the form of long-term funding of teams, rather than just projects, moving away from short-term R&D funding to the sustained financial support of successful organisations, programmes and services. While government funding could compromise trust in fact checkers or negatively influence perceptions about their independence in a UK context, research council funding can play a part in establishing the impact of fact checking and identifying areas for improvement. This helps to ensure the longevity of the fact checking ecosystem and to ensure that we are not constantly reinventing the wheel as a profession.

Equip fact checking organisations with tools that are fit for purpose

The changing information landscape produced by generative AI is creating an even more challenging environment for fact checkers. It is vital that they have access to the best tools, information and support from funders, online platforms, search engines and others to carry out their role effectively.

[127] Available at: Transparency Centre, *Reports Archive* (website), https://disinfocode.eu/reports-archive/?years=2024 (accessed 21 March 2024).

It is critical that while sophisticated detection tools are available for fact checkers, they are not available to all actors. It is important that anyone wishing to create harmful content is not able to take advantage of public access to such tools in order to hone their techniques and evade detection. As Sam Gregory of WITNESS puts it: “There’s a trade-off between security and access, which means if we make them available to anyone, they become useless to everybody.”[128]

Appropriate recipients of such tools include fact checking organisations that have signed up to public standards and principles and have been independently verified, for example by the European Fact-Checking Standards Network or the International Fact-Checking Network (IFCN). It is incumbent on anyone—whether technology companies, academics or others—to provide fact checkers with direct access to these tools if they can do so. If not, the government should introduce regulation and incentives so that those developing, contracting or acquiring monitoring and detection software are obliged to give access to less well-resourced frontline organisations.

These tools will provide fact checkers with increased speed and capacity, but it is also important to note that fact checkers will never rely solely on any automatic detection assessment when trying to establish whether a piece of content qualifies as misinformation or disinformation. The tools can only make an assessment of probability about whether content is AI-generated or manipulated, often to a high degree of accuracy, but it is always necessary for human beings to add context and caveat.

Support programmes that offer structured services for fact checkers

Such is the scale of the challenge that fact checkers cannot work alone, and they will rely increasingly on the development of a structured and reliable system of support. One positive example of a new programme is the Deepfakes Rapid Response Force set up by WITNESS. The pilot, established in spring 2023, saw a brokered service linking fact checkers in the IFCN with media forensics experts in academia and private sector companies that have models for detection.

When fact checkers encounter a potential deepfake, the WITNESS team triage—often discounting shallow fakes, which they define as “mis-contextualization, misattribution, or simple editing of video and audio”[129]—and escalate cases to the experts, working on the basis that an initial answer is required within hours, and a longer explanation within days.

[128] Sam Gregory, ‘When AI can fake reality, who can you trust?’, TED Democracy, November 2023, https://www.ted.com/talks/sam_gregory_when_ai_can_fake_reality_who_can_you_trust.

[129] UK Parliament, ‘WITNESS evidence to the House of Lords Communications and Digital Select Committee inquiry: Large language models’, p.8, 5 September 2023, https://committees.parliament.uk/writtenevidence/124270/pdf.

The service is intended to be used in critical human rights situations, including elections, and contexts where the threat of violence exists.[130] [131]

There is a growing need for a set of bespoke services of this kind to be in place nationally and internationally so that fact checkers and others do not waste time and money when needing to source analysis and forensic services, especially as the market for detection tools is expanding rapidly and the quality of services varies widely. It is important to establish the independence of such systems in order to generate public trust, and this may require regulatory intervention.

Although they shouldn't be relied upon to create the kind of services we describe here, commercial actors, including social media, search engine and AI companies, do have a key role to play in supporting the work of public interest organisations seeking to verify and fact check content and claims. While companies are the ones that have the tools and computing power, society needs trusted independent entities like fact checking organisations to provide the public with a service focussed on assessing content and claims.

Remove unacceptable barriers many online platforms and search engines have placed on access to tools

The mass popularity of generative AI, and the risks which accompany it, has emerged just as other tools that helped equip fact checkers, journalists and researchers understand and combat harmful misinformation and disinformation are being taken out of reach or made less useful. For example:

- Changes at X/Twitter around its API[132] have put access for fact checkers out of reach. This was previously a valuable way of identifying what kind of bad information was spreading and which actors might be involved. Many organisations and researchers suspect they have been excluded because of what they might find out,[133] and the European Commission has opened formal proceedings to assess whether X has breached the Digital Services Act (DSA) regarding data access for researchers, among other issues.[134]

[130] Rest of World, 'An Indian politician says scandalous audio clips are AI deepfakes. We had them tested', 5 July 2023, https://restofworld.org/2023/indian-politician-leaked-audio-ai-deepfake/.

[131] Sam Gregory, 'When AI can fake reality, who can you trust?', TED Democracy, November 2023, https://www.ted.com/talks/sam_gregory_when_ai_can_fake_reality_who_can_you_trust.

[132] The Verge, 'Twitter just closed the book on academic research', 31 May 2023, https://www.theverge.com/2023/5/31/23739084/twitter-elon-musk-api-policy-chilling-academic-research.

[133] The Guardian, 'Elon Musk threatens to sue Anti-Defamation League over lost X revenue', 5 September 2023, https://www.theguardian.com/technology/2023/sep/05/elon-musk-sue-adl-x-twitter.

[134] European Commission, 'Commission opens formal proceedings against X under the Digital Services Act', 18 December 2023, https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6709.

- Facebook's CrowdTangle was, for many years, the go-to tool for fact checkers, journalists and researchers to monitor misinformation and disinformation on the platform. It surfaced misleading claims to fact check and provided insight into Facebook moderation and policy implementation. Facebook has reduced the resources to sustain CrowdTangle and moved to phase it out.[135]

In an announcement in February 2024 about making more data from its platforms available to academic researchers, Meta revealed that its Third Party Fact-Checking partners would have access to its Content Library "to help them investigate and debunk misinformation".[136] This timing coincided with the deadline that such platforms were given to comply with the DSA.[137] It is currently unclear if this will replace all functionality currently available via CrowdTangle, and Meta have confirmed that the existing service will be permanently shut down on the 14th August 2024.[138] This is deeply concerning in a year when elections are taking place around the world.

There are some examples of good partnerships and structured collaboration between fact checkers and the online platforms and search engines they monitor, such as Meta's Third-Party Fact-Checking[139] or the Elections 24 project in Europe, which is supported by the Google News Initiative.[140]

But such examples are few and far between, and serious concerns arise even in areas where legislation and incentives exist. A recent report from the European Fact-Checking Standards Network (EFCSN)[141] showed that, even as the implementation of the DSA entered its key stage, the major online platforms and search engines were not fulfilling the promises they made to support fact checking, which they voluntarily signed up to in the Code of Practice on Disinformation. This included important commitments to provide fact checkers with access to the data that they need to maximise the quality and impact of their work.

[135] The Verge, 'Meta reportedly plans to shut down CrowdTangle, its tool that tracks popular social media posts', 23 June 2022, https://www.theverge.com/2022/6/23/23180357/meta-crowdtangle-shut-down-facebook-misinformation-viral-news-tracker.

[136] Meta, 'New Tools to Support Independent Research', 21 November 2023, https://about.fb.com/news/2023/11/new-tools-to-support-independent-research/.

[137] European Commission, *The enforcement framework under the Digital Services Act* (website), https://digital-strategy.ec.europa.eu/en/policies/dsa-enforcement (accessed 21 March 2024).

[138] NiemanLab, ' A window into Facebook closes as Meta sets a date to shut down CrowdTangle', 14 March 2024, https://www.niemanlab.org/2024/03/a-window-into-facebook-closes-as-meta-sets-a-date-to-shut-down-crowdtangle/.

[139] Meta, *Meta's Third-Party Fact-Checking Program* (website), https://www.facebook.com/formedia/mjp/programs/third-party-fact-checking (accessed 21 March 2024).

[140] *Elections24Check* (website), https://elections24.efcsn.com/ (accessed 21 March 2024).

[141] European Fact-Checking Standards Network, 'EFCSN reviews big tech's implementation of the EU code of practice on disinformation', 24 January 2024, https://efcsn.com/cop-review/.

These voluntary commitments by the companies were part of the co-regulatory approach taken in building the DSA. But the EU has emphasised that in the case of systematic failure to comply with the codes of conduct, the Commission and the Board may invite signatories to take necessary action. The full implications of this approach are not yet clear, but they do suggest that online platforms may be held to account.

Compliance with commitments of the Code of Practice on Disinformation

Service	Agreements and fact-checking coverage	Integration and use of fact-checking	Access to information for fact-checkers
YouTube	🔴	🔴	🔴
Google Search	🟡	🟡	🔴
Facebook	🟢	🟢	🟡
Instagram	🟢	🟡	🟡
TikTok	🟡	🟡	🔴
WhatsApp	🟡	🟢	🟡
Bing	🔴	🟡	🔴
Linkedin	🔴	🔴	🟡
X - Twitter	🔴	🔴	🔴
Telegram	🔴	🔴	🔴

Accelerate effective data access needs in the UK

While the EU DSA does allow for researcher and civil society access to data from platforms and search engines under certain circumstances, the UK's Online Safety Act does not do the same. Full Fact was part of a coalition of researchers and campaigners with expertise in online harms and their real-world consequences that urged the UK Government to compel greater transparency from online platforms and search engines. The aim was to give independent researchers access to data that reveals what is taking place on regulated internet platforms in real time.[142] Unfortunately, only minor concessions were made on this, and the Online Safety Act requires only that Ofcom produce a report on data access within 18 months[143] i.e. by mid 2025, along with associated guidance.

[142] Center for Countering Digital Hate et al., joint letter to UK Government: Data Access in Online Safety Bill, 19 June 2023, https://counterhate.com/wp-content/uploads/2023/06/Coalition-letter-OSB-data-access-amend-13_06_23-3.pdf.

[143] Online Safety Act 2023, ch. 50, section 162, https://www.legislation.gov.uk/ukpga/2023/50/section/162/enacted.

This is not enough. Full Fact will continue to press for greater transparency from companies so that we understand the impact of the choices they make, especially around harmful misleading information. It is clear from our experience that transparency and access to data will not be sufficiently forthcoming without an effective regulatory regime. This leaves a huge gap in terms of understanding the effectiveness of labelling AI content, or of computational detection models and what data has been used to train them. The research community around fact checking would benefit hugely from access to such data.

The Data Protection and Digital Information Bill has been amended to include a clause on access to data for vetted researchers.[144] Much of this amendment is welcome, for example ensuring that data can be accessed through an online database or API. But the amendment does not create sufficient change, because the description of eligible researchers envisaged is unlikely to include anti-disinformation researchers or journalists working at a fact checking charity or NGO. Therefore, this issue needs to be addressed through legislation in the next parliament.

Provide sustained funding to support the development of AI tools produced by fact checking organisations

Fact checking organisations have been using AI for many years to monitor and prioritise checkable claims, and continue to innovate in this space. Before the proliferation of consumer-accessible AI-generating tools, the rise of online misinformation and disinformation already presented a set of significant challenges for fact checkers across the world, not least about the scale of information available, the speed at which it is distributed, and the complexity of information incidents and crises such as elections and conflict.[145] Addressing misinformation at web scale requires purpose-built AI tools to make the work of human fact checkers more efficient and impactful.

Full Fact is a recognised leader in AI. Our tools are used by fact checkers around the world to find, check and challenge false claims, with 100,000 potential claims routinely identified each day.[146] In 2019, Full Fact won the international Google.org AI Impact Challenge which supported our work to use machine learning to improve and scale fact checking. It established cooperation with international experts to define how AI could transform this work, and to develop new tools for that purpose.

[144] UK Parliament, Data Protection and Digital Information Bill (session 2022-23, 2023-24), Lord Bethell's amendment, After Clause 27, available at: https://bills.parliament.uk/bills/3430/stages/18402/amendments/10011839 (accessed 21 March 2024).

[145] R. Llorente, 'Deepfakes in the Dock: Preparing International Justice for Generative AI',The SciTech Lawyer, Volume 20, Number 2, Winter 2024. https://www.gen-ai.witness.org/wp-content/uploads/2024/02/Deepfakes-in-the-Dock_Preparing-Intl-Justice-for-Generative-AI.pdf.

[146] Full Fact, 'How AI helps us detect 100,000 potential claims a day', 2 April 2021, https://fullfact.org/blog/2021/apr/ai-google-100000--claims-day/.

Full Fact AI is currently used in 20 countries and we are confident that it will continue to expand, especially with continued financial and technical development support from philanthropists, technology companies and others. It is important to recognise new technology as a huge opportunity as well as a threat, and celebrate the fact that AI tools employed effectively can help small teams of fact checkers deal with vast amounts of information and ensure public access to fact checks is timely and consequential.

Action for the government

- Consult with independent fact checking organisations, as well as academia and civil society working on related matters, to better understand the data access and tooling they need from the companies and wider experts that can provide it.
- Revisit the issue of ensuring that online platforms and search engines give verified civil society actors and researchers timely access to their data, so harmful misinformation and disinformation can be properly addressed.
- Ensure the timely introduction of any new regulatory provisions needed.

Action for online platforms and search engines

- Provide sustained funding and technical expertise to support the development of tools for fact checkers that help them spot AI-generated or edited content, whether these tools are developed internally or by others.
- Provide long-term funding for fact checking organisations and their networks rather than one-off grants.
- Ensure that structured dialogues are happening to share further expertise and insight with fact checkers.
- Share information on the reach and use of fact checks to tackle misinformation and disinformation on their platforms and consult with fact checkers when their work will be applied in new ways, for example via AI-powered chat assistants.
- Fulfil commitments made in the EU Code of Practice on Disinformation regarding fact checking, misinformation and disinformation.

Chapter 6: Government must provide resources for media literacy at the scale needed

Evaluation, funding and research can support citizens to navigate the new information environment

Recommendations: The government must increase resources for media literacy now and to meet future demand. Ofcom should work with online platforms and search engines to ensure that media literacy interventions are responding to the needs of UK citizens, and are seen by as many people as possible.

Effective media literacy provision in the UK hangs in the balance

Full Fact has argued before that good media literacy is the first line of defence against bad information online.[147] [148] Ofcom defines media literacy as “the ability to use, understand and create media and communications in a variety of contexts”.[149] It can be the difference between making decisions based on sound evidence, and making decisions based on poorly informed opinions. These can harm health and wellbeing, social cohesion, and democracy.

The guardians of the UK’s media literacy—the government, and the regulator Ofcom—need to make media literacy a priority. Both have an important role to play in rising to the challenge of improving media literacy now, as well as securing it for the future. This includes delivering targeted research, taking a creative and inclusive approach to ensuring a wide range of participants feed into media literacy needs and delivery, and dramatically increasing funding available for initiatives. As technology improves, media literacy programmes need to play a leading role in strengthening and adapting existing skills, as well as teaching new skills and knowledge to help citizens navigate the information environment.

[147] Full Fact, ‘Full Fact Report 2022’, ch.1, February 2022, https://fullfact.org/about/policy/reports/full-fact-report-2022/report/.

[148] Full Fact, ‘Full Fact Report 2023’, ch.11, March 2023, https://fullfact.org/about/policy/reports/full-fact-report-2023/report/.

[149] Ofcom, *Making Sense of Media* (website), https://www.ofcom.org.uk/research-and-data/media-literacy-research, (accessed 22 March 2024).

Media literacy needs to combine new knowledge and established skills

Research from Ofcom in 2023 into the online landscape in the UK found that misinformation was the most prevalent potential harm encountered by adults online, with two in five of them reporting having seen misinformation in a four-week period.[150] This included misinformation with political or electoral content, content which discriminated on the grounds of a protected characteristic, and financial and health misinformation. A nationally representative survey carried out by Ipsos UK and Full Fact in December 2023 indicated that one in four UK adults finds it difficult to distinguish true information from false information, and that one in three adults had falsely believed a news story was real until they found out it was fake.[151] Whilst not directly comparable, this appears to suggest that a large proportion of the population is seeing misinformation online, yet is not feeling confident about being able to tell whether something is true or false.

In an open society, media literacy must be a core part of the UK's defence against misinformation and disinformation. This needs to include regular evaluation and adaptation. Programmes need to keep up with the rapid evolution of technology as much as possible. Without this, the UK's collective media literacy curriculum could become outdated from one year to the next. For example, tips on spotting deepfakes—such as checking ears or the alignment of eyes—will become stale as the technology improves.[152] This has been recognised in Parliament:

> “The Online Safety Act will start to make modest progress towards media literacy, and people understanding and asking questions about factual accuracy and where something comes from when they see it on the web. It will go some way to addressing the first of the two sources of misinformation and disinformation—people telling lies, making stuff up, deepfakes of one kind or another. The sad fact is that the chances of deepfakes getting better with the advent of artificial intelligence is very high indeed so that, even if we think we can spot them now, we are probably kidding ourselves and in a year or two's time it will be doubly, trebly or quadruply difficult to work out what is real and what is completely made up.”
>
> **John Penrose MP during debate on “Online Filter Bubbles”[153]**

[150] Ofcom, ‘Online Nation 2023 Report‘, 28 November 2023, https://www.ofcom.org.uk/__data/assets/pdf_file/0029/272288/online-nation-2023-report.pdf.

[151] Ipsos, ‘Full Fact UK Public Attitudes Research’, April 2024, http://fullfact.org/audience-research-2023.

[152] Full Fact, ‘How to spot deepfake videos and AI audio’, 20 December 2023, https://fullfact.org/blog/2023/dec/how-to-spot-deepfakes/.

[153] House of Commons, Westminster Hall Debate: Online Filter Bubbles: Misinformation and Disinformation, 16 January 2024, vol. 743 Column 246WH, https://hansard.parliament.uk/Commons/2024-01-16/debates/9BA38505-4297-4CFC-A009-4A617BC682A9/OnlineFilterBubblesMisinformationAndDisinformation.

Part of the challenge is about raising levels of technical awareness: are users aware that they are interacting with generative AI? Do they know that generative AI only produces probable answers? Users need to combine this knowledge with more traditional media literacy skills such as critically evaluating and assessing the accuracy of information being presented to them. Seeking out cues from news media about how they are using technology to produce journalistic content will also be important.

Looking ahead to potential future developments is essential. As the House of Commons Public Accounts Committee warned in its recent report into the government's preparedness for Online Safety regulation, expectations are high, but "it may be years until people notice a difference to their, and their children's, online experience".[154]

Ofcom's recent discussion paper on generative AI and media literacy concludes that generative AI "does not necessarily mean completely new media literacy skills are needed".[155] The paper argues that "understanding and shaping how generative AI will change our world, and what it means for media literacy, will be an important task for the years ahead", and that many of the necessary skills are already required when navigating today's internet, and that higher levels of the same skills will be needed to apply these in different ways.

Ofcom's new duties to promote media literacy could build public resilience to misinformation and disinformation

In 2023, Full Fact successfully campaigned for an amendment to the Online Safety Bill that updated Ofcom's media literacy duty. This introduced new objectives to Ofcom's role, relating specifically to social media and search platforms. Ofcom is now required to help members of the public establish the reliability, accuracy and authenticity of information they encounter online, and to understand how to better protect themselves and others from misinformation and disinformation.[156]

Ofcom also has a duty to encourage the development and use of technologies and systems that support users of regulated services to protect themselves and others online, including on misinformation and disinformation. This may include providing users with further context about content they encounter, or signposting users to resources, tools or information which raises awareness about how to use regulated services. The overall aim must always be to mitigate the harms of misleading information.

[154] House of Commons Committee of Public Accounts, 'Preparedness for onlinesafety regulation - Thirteenth Report of Session 2023–24', 5 February 2024, https://committees.parliament.uk/publications/43321/documents/215761/default/.

[155] Ofcom, 'Future Technology and Media Literacy: Understanding Generative AI', 22 February 2024,https://www.ofcom.org.uk/__data/assets/pdf_file/0033/278349/future-tech-media-literacy-understanding-genAI.pdf.

[156] Full Fact, 'Full Fact campaign wins improved media literacy in the Online Safety Bill', 28 July 2023, https://fullfact.org/blog/2023/jul/full-fact-campaign-wins-improved-media-literacy-in-the-online-safety-bill/.

In spring 2024, Ofcom will consult on its new media literacy strategy and—following consultation—publish its first iteration of that strategy under the new duties. If implemented effectively, this will build the public's resilience to misinformation and equip citizens with the skills needed to recognise and act on online harms.

Ofcom's existing programme of work to help improve the online skills, knowledge and understanding of UK adults and children is called Making Sense of Media. To date, its work includes sharing evidence and research on UK citizens' media habits and attitudes, and working with the media literacy community to pilot new initiatives.[157]

This programme also has a network of 460 members whose purpose is to increase collaboration, information-sharing and debate, to improve media literacy in the UK.[158] Ofcom should continue to invite a variety of participants to this programme, including those who do not view themselves as part of the media literacy sector, but who can provide insight and evidence about harmful misinformation (such as misleading financial or health information), such as early years coordinating networks or health condition support groups.

Ofcom should ensure this mix includes technology educators. As highlighted above, media literacy over the next decade needs to include public education about what new technology can do and how it works. It should ensure that citizens are not intimidated by technological changes, and are well equipped to take advantage of them. It should also seek to inspire confidence that the right balance is being sought between protection from harms and freedom of expression.

Diversifying input to, and delivery of, media literacy programmes is also more likely to result in success, as a "one size fits all" approach will not work. Too often, policy discourse portrays media literacy as an impenetrable abstract concept, which is hard for many people to understand. In fact, media literacy initiatives in the UK are diverse.[159] They range from providing satellite-based internet to rural communities[160] and digital

[157] Ofcom, *Making Sense of Media* (website), https://www.ofcom.org.uk/research-and-data/media-literacy-research, (accessed 22 March 2024).

[158] Ofcom, *Join the Making Sense of Media Network* (website), https://www.ofcom.org.uk/research-and-data/media-literacy-research/network (accessed 22 March 2024).

[159] Ofcom, 'New Ofcom study explores how media literacy can support mental health', 15 May 2023, https://www.ofcom.org.uk/news-centre/2023/new-ofcom-study-explores-how-media-literacy-can-support-mental-health.

[160] F. Williams, L. Philip, J. Farrington, & G. Fairhurst, ''Digital by Default' and the 'hard to reach': Exploring solutions to digital exclusion in remote rural areas', Local Economy, 31(7), 757-777, 30 September 2016, https://doi.org/10.1177/0269094216670938.

peer support,[161] through to podcasts for parents[162] and workshops for educators.[163] Research by Full Fact into what makes an effective media literacy intervention found that good practice is "not confined to the structures of classrooms", and that even "brief training sessions of 15 minutes can improve media literacy to some extent".[164] This is echoed in a recent evaluation of the UK media literacy landscape commissioned by the Department for Science, Innovation and Technology (DSIT), which recommends embedding media literacy "in services people already use".[165]

Government funding should enable an ambitious Online Media Literacy Strategy

The government has already committed to publishing an annual Media Literacy Action Plan about initiatives to be delivered that forthcoming year, which are intended to respond to the needs of the media literacy sector.[166] The first Online Media Literacy Strategy,[167] published in 2021, intended "to improve national media literacy capabilities by supporting the media literacy sector to undertake activity in a more effective, wide-reaching, and coordinated way"[168]. There were some important conclusions about where to focus, including provisions for vulnerable users and engaging with hard-to-reach audiences, as well as filling gaps in evaluation and long-term stable funding for media literacy initiatives.

Unfortunately, these good ideas have not been matched with sufficient resources. The latest plan[169] published by DSIT sees just £2 million for 13 initiatives through grant funding. While the programming may be well focused and the research delivered valuable, it is not anywhere near ambitious enough to tackle either existing needs or future media literacy challenges, as identified in a recent evaluation by the London

[161] National Health Service, *Support Hope and Recovery/Resource Online Network (SHaRON)* (website), https://www.sharon.nhs.uk/ (accessed 22 March 2024).

[162] Internet Matters, *Fostering Digital Skills - Online learning course for foster carers* (website),https://www.internetmatters.org/fostering-digital-skills-online-learning-course/ (accessed 22 March 2024).

[163] Wise Kids, *Working with Educators* (website), https://wisekids.org.uk/wk/for-educators/ (accessed 22 March 2024).

[164] Full Fact, 'Media and information literacy: Lessons from interventions around the world', February 2020, https://fullfact.org/media/uploads/media-information-literacy-lessons.pdf.

[165] L. Edwards, V. Obia, E.Goodman & S. Spasenoska, 'Cross-sectoral challenges to media literacy - Final Report', UK Government Department of Science Innovation and Technology, August 2023, https://assets.publishing.service.gov.uk/media/651167fabf7c1a0011bb4660/cross-sectoral_challenges_to_media_literacy.pdf.

[166] UK Government, 'Online Media Literacy Strategy', 14 July 2021, https://www.gov.uk/government/publications/online-media-literacy-strategy.

[167] UK Government, 'Online Media Literacy Strategy', 14 July 2021, https://www.gov.uk/government/publications/online-media-literacy-strategy.

[168] UK Government, 'Year 3 Media Literacy Action Plan (2023/24)', 23 October 2023, https://www.gov.uk/government/publications/year-3-media-literacy-action-plan-202324.

[169] UK Government, 'Year 3 Media Literacy Action Plan (2023/24)', 23 October 2023, https://www.gov.uk/government/publications/year-3-media-literacy-action-plan-202324.

School of Economics.[170] [171] DSIT must make the case for a significant increase in funding dedicated to promoting media literacy in its forthcoming annual plan, likely to be published in summer 2024.

The 2019 Conservative manifesto did not explicitly include media literacy in its intent to make the UK "the safest place in the world to go online". This meant that other departments were not obliged to focus on online safety through the lens of media literacy. For example, there was no team working on this within the Department for Education. Parties should not make the same mistake this time. Manifestos must contain specific and tangible commitment to improve media literacy, ideally within their education agenda. This will help to avoid repeating the lack of cross-departmental coherence under the current government. Ultimately, the Cabinet Office should coordinate a media literacy agenda across Whitehall, with activity taking place within the Department for Education, and with DSIT supporting online safety and the Department for Culture, Media and Sport (DCMS) supporting the future of journalism.

As the election approaches, and the media and civil society organisations consider how to work together to address misinformation and disinformation, another opportunity has emerged: prebunking. Prebunking covers both the active provision of information about election hot topics, for example by national statistics institutes, as well as the process of "debunking lies, tactics or sources before they strike"[172] which may be better undertaken by civil society organisations.

Independent fact checkers like Full Fact are well placed to proactively identify false and misleading claims or tactics which may arise, and to work with others like the Electoral Commission, media, online platforms and search engines in order to warn audiences in advance and help inoculate them from harm. This should be a public priority in the run-up to the general election, and should be welcomed by all political parties who care about an honest and accurate election campaign.

Online platforms and search engines must help to educate users

What happens on a platform at point of use really matters. It could be a notification about the source of the information, a prompt to consider if you really want to post something, or a fact check about the content. Ofcom research shows that interventions

[170] L. Edwards, V. Obia, E.Goodman & S. Spasenoska, 'Cross-sectoral challenges to media literacy - Final Report', UK Government Department of Science Innovation and Technology, August 2023, https://assets.publishing.service.gov.uk/media/651167fabf7c1a0011bb4660/cross-sectoral_challenges_to_media_literacy.pdf.

[171] UK Government, 'Media literacy uptake amongst 'hard to reach' citizens', 29 September 2023, https://www.gov.uk/government/publications/media-literacy-uptake-amongst-hard-to-reach-citizens.

[172] First Draft, 'A guide to prebunking: a promising way to inoculate against misinformation', 29 June 2021, https://firstdraftnews.org/articles/a-guide-to-prebunking-a-promising-way-to-inoculate-against-misinformation/.

into consumption of misinformation are valued by users: "Overlays and labels about misinformation were considered the most useful by the participants who encountered them, as they appreciated being warned about potentially upsetting or misleading content. Some participants thought that these interventions showed that the platform 'cared' about its users and was trying to provide them with a good user experience."[173] Yet some companies continue to avoid labelling or contextualising information that people are consuming.

Meta, Google and TikTok, are active in Ofcom's "media literacy by design" work,[174] which has the potential to succeed if product teams get behind the effort. Since the Online Safety Act places no requirements on online platforms and search engines to undertake media literacy initiatives for their users, Ofcom should combine the wider transparency reporting described in Chapter 1 with its user research powers, to bridge the gap between what users find useful and what platforms are willing to share.

It is vitally important that online platforms and search engines treat this as a priority, and future versions of online safety legislation must ensure that the largest platforms are given a duty to provide media literacy programmes which meet users' needs.

Action for Ofcom

- Recommend how online platforms and search engines can advance media literacy, including helping users to judge information and slow its spread.
- Ahead of the next election, publish a plan specifically about risks and mitigation of misinformation and disinformation during the election, including how Ofcom would respond to a request by a Secretary of State to prioritise media literacy during an election information incident (see Chapter 7).

Action for the government

- Ahead of the next election, increase Ofcom's spending cap and issue a grant specifically to fund Ofcom's delivery of its media literacy duty, opening up real resources to make change.
- Significantly increase funding dedicated to civil society and media initiatives that promote media literacy.
- In any future legislation on online harms, establish a duty for online platforms and search engines to provide media literacy programmes which meet users' needs now and in future.

[173] YouGov, 'User Attitudes towards On-Platform Interventions', 30 October 2023,https://www.ofcom.org.uk/__data/assets/pdf_file/0020/270371/ofcom-interventions-qual-report.pdf.

[174] Ofcom, *Our Establish Working Group* (website), https://www.ofcom.org.uk/research-and-data/media-literacy-research/approach/establish (accessed 22 March 2024).

Action for political parties

- Ahead of the next parliament, all parties should make manifesto commitments to improve media and digital literacy to allow the next generation to make the most of everything the internet can offer while keeping safe.

Chapter 7: Protect democracy from misinformation and disinformation in the age of AI

The UK needs transparency and better planning for information incidents to protect future elections

Recommendation: The government should set out how it will work transparently with online platforms and search engines to challenge misinformation and disinformation during the next general election, including in the event of an information incident.

Likely risks for the UK election include convincing fakes and denial of information which is true

There has been widespread concern among parliamentarians[175] [176], the Electoral Commission[177] and civil society groups about the UK's democratic process being increasingly vulnerable to misinformation and disinformation as easy-to-use AI tools become available on demand.

It is unlikely that any single piece of deepfake video or audio will end up swaying significant numbers of views or votes in an election in the UK. Even the efficacy of the notorious deepfake released two days before the Slovakian election, which created a fake conversation between a journalist and a leading politician, is not clear to political experts in the country.[178] But high volumes of false content and sophisticated campaigns and hoaxes could threaten to swamp online platforms and would be difficult to debunk just before polling day, especially when no broadcast coverage of campaigning or election issues is allowed while polls are open.

[175] The Guardian, 'Call for action on deepfakes as fears grow among MPs over election threat', 21 January 2024, https://www.theguardian.com/politics/2024/jan/21/call-for-action-on-deepfakes-as-fears-grow-among-mps-over-election-threat.

[176] House of Commons debate 'Political Parties, Elections and Referendums', 31 January 2024, vol. 743 column 903, https://hansard.parliament.uk/Commons/2024-01-31/debates/99BAC8DE-6003-4473-8B92-CD628AA6D859/PoliticalPartiesElectionsAndReferendums.

[177] The Guardian, 'Time running out for UK electoral system to keep up with AI, say regulators', 28 June 2023, https://www.theguardian.com/politics/2023/jun/28/time-running-out-for-uk-electoral-system-to-keep-up-with-ai.

[178] The Times, 'Was Slovakia election the first swung by deepfakes?', 7 October 2023, https://www.thetimes.co.uk/article/was-slovakia-election-the-first-swung-by-deepfakes-7t8dbfl9b.

According to Ipsos UK research commissioned by Full Fact, three out of four members of the public think misinformation will have at least some impact on the general election result, with 54% believing generative AI will have an impact, and 58% concerned about the impact of deepfakes.[179]

The emergence of generative AI and especially large language models will be a new element at the next election. While the technology is moving at a pace that makes it hard to predict exactly what impact it might have, research is emerging on the potential risks. A study conducted by OpenAI, in collaboration with researchers from Georgetown and Stanford universities, notes that "for malicious actors looking to spread propaganda—information designed to shape perceptions to further an actor's interest—these language models bring the promise of automating the creation of convincing and misleading text for use in influence operations, rather than having to rely on human labour".[180]

This framing, and the scale of the threat it implies, forms a significant part of the debate in the UK, which often concentrates on potential foreign interference, including through the use of AI-produced disinformation created by malign foreign actors.[181]

However, cheap and powerful tools are now widely available to anyone who wants to create mischief, or simply amuse themselves or others. This is why it is so challenging to make an accurate assessment of the real threat and where it will emerge. The information environment may become so polluted with false content that it becomes difficult to know what you can trust to be true. Former Justice Secretary Sir Robert Buckland is among those who have warned about a situation where truth can be easily denied because so many fakes exist. Full Fact has not yet seen examples of this in the UK, but complacency around what may happen should be avoided. With trust in politicians already at a 40-year low,[182] care must be taken not to further damage public confidence in our democracy.

Sir Robert told the BBC about another concern: "Those who want to undermine the [electoral] process will simply say attempts to deal with deepfakes are censorships rather than something more legitimate designed to protect the sanctity of the truth".[183] And as the Chair of the Electoral Commission John Pullinger has pointed out,[184] giving more attention to deepfakes could divert media coverage from the real campaign.

[179] Ipsos, 'Full Fact UK Public Attitudes Research', April 2024, http://fullfact.org/audience-research-2023.

[180] J. Goldstein, G. Sastry, M. Musser, R. DiResta, M. Gentzel, K. Sedova, 'Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations', OpenAI, January 2023, https://cdn.openai.com/papers/forecasting-misuse.pdf. See also a topline summary on the OpenAI website here: https://openai.com/research/forecasting-misuse (accessed 22 March 2024).

[181] National Cyber Security Centre, 'NCSC Annual Review 2023', 14 November 2023, https://www.ncsc.gov.uk/collection/annual-review-2023/resilience/case-study-defending-democracy.

[182] IPSOS, 'Trust in politicians reaches its lowest score in 40 years', 14 December 2023, https://www.ipsos.com/en-uk/ipsos-trust-in-professions-veracity-index-2023.

[183] BBC News, 'Fears UK not ready for deepfake general election', 21 December 2023, https://www.bbc.co.uk/news/uk-politics-67518511.

[184] Financial Times, 'Voter ID rules could be seen to benefit Tories, says UK elections watchdog', 22 January 2024, https://www.ft.com/content/7db6f3f7-d1e7-4c2d-871b-b93e14abf0dd.

Legislation is not sufficient to protect individuals and democracy from misinformation and disinformation during an election campaign

These concerns about the potential erosion of trust in democratic processes are urgent. But so far the government's legislative response has been limited.

As one of the organisations which pressed for their introduction, Full Fact welcomes the new requirement for digital campaigning material to display a digital imprint in the next election through the Elections Act. This will provide information on who has published the political or campaign material. However, this UK law is unlikely to deter those who are attempting to influence public opinion on behalf of another state.

For that, we need to turn to the new foreign interference offence set out in the National Security Act 2023[185]. This came into effect on 20 December 2023 and is included as a priority offence in the Online Safety Act.

The government says that the "principal aim" of this offence is "to create a more challenging operating environment for, and to deter and disrupt the activities of, foreign states who seek to undermine UK interests, our institutions, political system, or our rights, and ultimately prejudice our national security"[186]. It has tried to position the offence as something that will make a difference in the fight against online disinformation, particularly during elections.[187] [188] But Full Fact is sceptical about the extent to which it will be effective during the upcoming campaign.

The government says the priority offence under the Online Safety Act "will require digital platforms to proactively take action against a wide range of state-sponsored disinformation" including "digitally manipulated content where this has the aim of interfering with UK elections" and "where this is AI-generated". But there is no new disinformation offence per se and there are no new electoral offences. Only if three conditions are met[189], involving intent, illegitimacy and the participation of a foreign power, could an individual end up with a tougher sentence than has previously been the case for election offences.

[185] National Security Act 2023, ch.32, Section 13, https://www.legislation.gov.uk/ukpga/2023/32/part/1/crossheading/foreign-interference/enacted.

[186] UK Government, 'Foreign interference: National Security Bill factsheet', 12 February 2024, https://www.gov.uk/government/publications/national-security-bill-factsheets/foreign-interference-national-security-bill-factsheet.

[187] UK Government, 'Foreign interference: National Security Bill factsheet', 12 February 2024, https://www.gov.uk/government/publications/national-security-bill-factsheets/foreign-interference-national-security-bill-factsheet.

[188] UK Parliament, written question: 'Misinformation: General Elections', UIN 185058, tabled on 15 May 2023, https://questions-statements.parliament.uk/written-questions/detail/2023-05-15/185058/.

[189] UK Government, 'Foreign interference: National Security Bill factsheet', 12 February 2024, https://www.gov.uk/government/publications/national-security-bill-factsheets/foreign-interference-national-security-bill-factsheet.

There are challenges in applying this. Where the offence is viewed as priority illegal content in the Online Safety Act, it is difficult to see how platforms can be expected to make the necessary judgments about an individual's behaviour and intent, in order to establish whether the content constitutes such an offence. It is also extremely challenging to identify coordinated networks in the midst of a campaign, and while large platforms do have some capabilities, the involvement of a foreign state may not be clear—particularly given that such involvement may well be indirect.

Attempts to regulate this comprehensively could easily lead to excessive moderation of content and breach the need for regulated services to protect freedom of expression.

Much of the responsibility for interpreting how this might work in practice online rests with Ofcom. But the foreign interference offence lacks a body of case law or academic discussion which the regulator can draw upon.[190] Ofcom acknowledges that it is "likely to be particularly difficult to identify in practice, because [it depends] heavily on context and on circumstances offline".

For the time being, Ofcom's proposed approach is that "our guidance should describe the offences and the questions a service should ask itself", adding "bots play an important role in generating and spreading content which is likely to amount to a foreign interference offence and we propose also to draw attention to this in our guidance".

Under the Online Safety Act the offence is, in many respects, still a work in progress. It formed part of Ofcom's Illegal Harms Consultation which closed in February 2024[191], and the next expected step is that Ofcom will issue its guidance to support social media platforms and search services in understanding their regulatory obligations when making judgements about the foreign interference offence. It is an open question whether the offence can work in any meaningful way at internet scale in the longer term, and it will not play a significant role this year. Ofcom's CEO Dame Melanie Dawes and Group Director for Online Safety Gillian Whitehead told Full Fact in January that the online safety provisions relating to elections will not be in place until after the next general election, and that they have informed DSIT that this is the case.

End unnecessary secrecy in government to tackle false information

Concerns about freedom of expression have rightly been at the forefront of public debate about the regulation of online harms. Full Fact has previously made proposals about

[190] Ofcom, 'Assessing the risk of foreign influence in UK search results', 19 September 2023, https://www.ofcom.org.uk/research-and-data/online-research/online-safety-research/assessing-the-risk-of-foreign-influence-in-uk-search-results.

[191] Ofcom, 'Protecting people from illegal harms online', Annex 10 p.135, 9 November 2023, https://www.ofcom.org.uk/__data/assets/pdf_file/0025/271168/annex-10-illegal-harms-consultation.pdf.

how to better protect freedom of expression online[192] [193]. We first raised concerns in the 2022 Full Fact report[194] about the government's reluctance to tackle misinformation and disinformation publicly through legislation, instead seeking to limit speech online by lobbying online platforms and search engines behind closed doors. Such an approach, amounting to censorship-by-proxy, could have been seen as an imperfect but understandable response to the emergency of the pandemic on a temporary basis, but it is not sustainable.

We had hoped that a more open and transparent system would emerge during the passage of the Online Safety Bill.[195] But in May 2023, the government admitted that it continues to meet regularly in private with social media platforms, to "aid our understanding of the spread of misinformation and disinformation on their services, including artificially manipulated media, and the range of steps they are taking to address this".[196] This includes attempting to influence terms of service, policies and enforcement mechanisms "whilst still respecting freedom of expression." And in November 2023, the Minister for Tech and the Digital Economy Saqib Bhatti told Parliament that DSIT would be "working closely with social media platforms to ensure that the right systems are in place to identify and remove harmful material, including deepfakes, where it breaches platform terms of service."[197] This entire process needs to be more transparent.

We have also expressed concerns about the lack of effective parliamentary oversight of the government's work monitoring and countering false information through the National Security Online Information Team (NSOIT), previously known as the Counter Disinformation Unit (CDU).[198] While the government has said that NSOIT is "focused exclusively on risks to national security and public safety", it is understood to include elections. Parliamentary scrutiny of this work has been repeatedly deflected in an unjustifiably secretive way, a consistent concern since the pandemic.

[192] Full Fact, 'Full Fact Report 2023', ch.9, March 2023, https://fullfact.org/about/policy/reports/full-fact-report-2023.

[193] Full Fact, 'Full Fact Report 2022', ch.7, February 2022, https://fullfact.org/about/policy/reports/full-fact-report-2022/report.

[194] Full Fact, 'Full Fact Report 2022', ch.7, February 2022, https://fullfact.org/about/policy/reports/full-fact-report-2022/report.

[195] Full Fact, 'Full Fact Report 2022', February 2022, https://fullfact.org/about/policy/reports/full-fact-report-2022/report/.

[196] UK Parliament, House of Commons written question: 'Misinformation: General Elections', UIN 185058, tabled on 15 May 2023, https://questions-statements.parliament.uk/written-questions/detail/2023-05-15/185058/.

[197] UK Parliament, House of Commons debate, 'AI-generated Content: Social Media', volume 740, col. 641, 15 November 2023, https://hansard.parliament.uk/Commons/2023-11-15/debates/926E3CE7-F123-4602-889A-161A6BF7394C/AI-GeneratedContentSocialMedia.

[198] UK Parliament, House of Commons written question: 'National Security Online Information Team', UIN 43, tabled on 7 November 2023, https://questions-statements.parliament.uk/written-questions/detail/2023-11-07/43/.

During a Lords debate on fake news in 2020, one parliamentarian said: "We have heard very little about its work and received no detail on what its achievements or actions are", and was promised a Ministerial statement to come "when time allows", with the government saying its "real focus" was to "act as expeditiously as possible." In response to a written question in 2021 to the Department for Digital, Culture, Media and Sport, the government refused to give basic factual information about how many anti-vaccination posts the CDU had reported to online platforms and search engines.[199]

It is breathtaking that the government continues to assume public support for such a secretive and unstructured approach to regulating online speech four years after the pandemic began. Despite warnings from Full Fact, and the opportunity to correct this situation through the Online Safety Act, the government continues to leave itself wide open to the accusation of extrajudicial state censorship, and has given the public no reasons to trust the effectiveness or justifiability of its approach so far.

There may be some necessary limits to transparency. Revealing some tactics may advantage bad-faith actors, for example. But more transparency about and oversight of NSOIT's activities—and other relevant initiatives, like the Defending Democracy Taskforce—would help to build trust in this work. Unnecessary secrecy around government attempts to counter false information should be brought to an end in future legislation or through changes to the Online Safety Act.

Security of elections: key government institutions

In addition to the security services (including the National Cyber Security Centre), the following government organisations have responsibilities around the security of elections, including misinformation and disinformation:

The Defending Democracy Taskforce This was established in November 2022 as a cross-departmental and inter-agency initiative "seeking to protect the democratic integrity of the UK from foreign influence" that includes the wider election security capability.[200] Its first meeting was chaired by security minister, Tom Tugendhat, and it reports into the National Security Council. Parliament's Joint Committee on National

[199] UK Parliament, House of Commons written question: 'Vaccination: Disinformation', UIN 90926, tabled on 10 December 2021, https://questions-statements.parliament.uk/written-questions/detail/2021-12-10/90926.

[200] The Defending Democracy Taskforce reports into the National Security Council (NSC) chaired by the prime minister. It is described as "a cross-departmental and inter-agency initiative made up of ministers and officials from policy-owning departments, including the Cabinet Office, Home Office, DSIT, DLUHC and DfE, law enforcement, the UK intelligence community and Parliament". See: UK Parliament, House of Commons written question: 'Defending Democracy Taskforce', UIN 182673, tabled on 25 April 2023, https://questions-statements.parliament.uk/written-questions/detail/2023-04-25/182673/.

Security Strategy (JCNSS) is currently conducting an inquiry on “defending democracy”[201] ahead of the General Election which includes some specific questions about the Taskforce. These cover its objectives, working methods, resources and achievements; what more it could be doing; and how its work informs “decisions of the National Security Council, the National Security Risk Assessment process and wider Government activity to counter state threats”.

The Joint Election Security Preparedness unit (JESP) Established by the Defending Democracy Taskforce, this takes overall responsibility for coordinating electoral security and driving the government’s election preparedness. It is a joint endeavour between the Elections Directorate in the Department for Levelling Up, Housing and Communities and the Government Security Unit in the Cabinet Office. The unit’s role is to identify and mitigate election security risks and to improve preparedness, as well as “horizon-scanning to get ahead of emerging risks and technological developments”. It works with those delivering the election, the security and intelligence community, and other government departments.

The National Security Online Information Team (NSOIT) This is a unit within the Department for Science, Innovation and Technology, previously known as the Counter Disinformation Unit (CDU) and sometimes the Rapid Response Unit.[202] The government has given undertakings that NSOIT is “focused exclusively on risks to national security and public safety”, understood to include elections. If it identifies content “which is assessed to breach the terms and conditions of the relevant platform it may share that content with the platform”. The platform then determines “whether or not to take any action in line with their terms of service”.

Tackling misinformation and disinformation must be done in public

Other approaches can help protect our democracy from online harms. In the 2022 and 2023 Full Fact reports, we set out the need for a protocol to warn the public about threats identified by security services during an election campaign.[203] Canada has a protocol for such situations, but the UK does not.[204] With the widespread introduction of

[201] UK Parliament National Security Strategy Joint Committee, ‘JCNSS launches inquiry on Defending Democracy with UK election expected this year’, 1 February 2024, https://committees.parliament.uk/committee/111/national-security-strategy-joint-committee/news/199739/jcnss-launches-inquiry-on-defending-democracy-with-uk-election-expected-this-year.

[202] UK Parliament, House of Commons written question: ‘National Security Online Information Team’, UIN 43, tabled on 7 November 2023, https://questions-statements.parliament.uk/written-questions/detail/2023-11-07/43/.

[203] Full Fact, ‘Full Fact Report 2023’, March 2023, https://fullfact.org/about/policy/reports/full-fact-report-2023.

[204] Government of Canada, *Critical Election Incident Public Protocol* (website), https://www.canada.ca/en/democratic-institutions/services/protecting-democracy/critical-election-incident-public-protocol.html (accessed 22 March 2024).

consumer AI tools, this leaves our electoral systems and processes more vulnerable to misinformation and disinformation than ever.

Having a protocol enables a non-partisan determination of whether to inform the public that an incident that threatens the integrity of an election has arisen. Unless there are concerns about unnecessary amplification of the incident, the public can be informed about it and any steps they should take to protect themselves. Canada's model has been independently assessed and could be adapted for the UK.[205] We repeat our recommendation from the 2023 Full Fact report that the Minister for the Cabinet Office, who has responsibility for both defending democracy and for electoral law, should initiate a process to bring about a UK Critical Election Incident Public Protocol through non-legislative means.

More broadly, parties should set out their policies for protecting our electoral processes from misinformation and disinformation in their manifestos.

In government, the Conservatives have not yet set out such policies despite opportunities to do so. There is little in the AI white paper on this, for example.[206] In response to a written question in January 2024 on AI deepfakes during elections, security minister Tom Tugendhat said simply there is "a robust system to rapidly respond to any threats during election periods".[207] This does nothing to dispel our concerns about the lack of transparency from the government in this area. The answer also referred to a discussion about AI threats to democracy during the AI Safety Summit in November 2023.[208] Given that there is a rapidly decreasing timeframe during which those threats can be mitigated prior to the UK election, the government should indicate what progress has been made on developing a shared understanding of misinformation and disinformation risks, as stated at the Summit.

Shadow Foreign Secretary David Lammy has said that the Labour Party is committed to introducing regulation of companies developing the most powerful frontier AI, which could be used to disrupt elections. He cited concerns "on the use of AI and deepfakes to seed false narratives, spread lies and foment divisions'' including "the use of AI and widespread disinformation, misinformation and malinformation which undermines

[205] Government of Canada, 'Report on the assessment of the Critical Election Incident Public Protocol', 20 November 2020, https://www.canada.ca/en/democratic-institutions/services/reports/report-assessment-critical-election-incident-public-protocol.html.

[206] UK Government, 'AI regulation: a pro-innovation approach', 29 March 2023, https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach.

[207] UK Parliament, written question: 'Elections: Disinformation', UIN 11373, tabled on 24 January 2024, https://questions-statements.parliament.uk/written-questions/detail/2024-01-24/11373/.

[208] UK Government, 'AI Safety Summit 2023: Roundtable Chairs' Summaries', 3 November 2023, https://www.gov.uk/government/publications/ai-safety-summit-2023-roundtable-chairs-summaries-2-november/ai-safety-summit-2023-roundtable-chairs-summaries-2-november.

democracy."[209] Further detail of Labour's plans is not yet available.

Protecting our democracy must be done independently and openly

In the 2022[210] and 2023[211] Full Fact reports, we highlighted the urgent need for better protection of our electoral processes given the increasing threats posed by a highly connected online environment in which election misinformation and disinformation can spread rapidly and at scale. We said doing this effectively would require better regulation as well as collaborative responses from regulators, technology platforms and civil society. The emergence of generative AI has made the need even more pressing.

There is still time to protect our democracy and freedom of speech before the next general election. This must be done transparently and independently, including by introducing an election incident protocol, setting out what voters should expect to see from government and regulators during an election information incident, and ending government attempts to limit speech online without oversight. The government should think beyond the foreign interference offence and provide information openly so that its existing counter-disinformation initiatives have a better chance of becoming trusted.

Action for the government

- The Minister for the Cabinet Office should initiate a process to bring about a UK Critical Election Incident Public Protocol through non-legislative means to secure public confidence in how elections are protected.
- Consult more transparently with civil society organisations, with regulators and with internet platforms and search engines, to improve the latter's policies on supporting election integrity in the UK.
- Clarify how misinformation and disinformation will be challenged during and around the election through further law, regulation and non-legislative approaches.

Action for the Electoral Commission and Ofcom

- The Electoral Commission should monitor and assess the effectiveness of the new digital imprints regime, and identify any further improvements that may be needed to ensure greater transparency around electronic material.

[209] UK Parliament, House of Commons debate: 'Cyber Interference: UK Democracy', volume 742, col. 488, 7 December 2023, https://hansard.parliament.uk/Commons/2023-12-07/debates/D3F22078-B63B-4279-A459-97CE8738AD81/details.

[210] Full Fact, 'Full Fact Report 2022', ch.9, February 2022, https://fullfact.org/about/policy/reports/full-fact-report-2022/report/.

[211] Full Fact, 'Full Fact Report 2023', ch.8, March 2023, https://fullfact.org/about/policy/reports/full-fact-report-2023/report/.

- Ofcom should publish recommendations for stakeholders, including online platforms and search engines, on media literacy and UK elections, under its new media literacy duties on misinformation, as soon as possible ahead of the general election.

Chapter 8: Political parties using generative AI in campaigning must do so transparently and responsibly

Party leaders should give reasons for the public to trust what they are doing with generative AI to win our votes

Recommendation: Political parties should commit publicly to transparent and responsible use of AI during elections.

The need for transparent use of generative AI in political campaigning

Generative AI has many potential uses for parties in their campaigns, such as improving the speed of their work, testing through virtual focus groups to generate immediate feedback, and chatbots to intervene in social media channels.[212] [213] Some use cases might be seen as more or less acceptable by the public or by other parties. For example, recent polling suggested that the public are wary of trusting AI to help make subjective choices about who to vote for, but more willing to consider the use of AI in helping to find useful information such as how to register to vote.[214]

Generative AI is already being used around the world in an effort to gain votes, with varying levels of honesty and disclosure.

In April 2023 the US Republican National Committee (RNC) used AI images in an advert to present a dystopian vision of what it argues a future term of President Joe Biden would look like if he were re-elected. The images were disclosed as being AI-generated.[215]

[212] See, for example those listed in: Prospect, 'The dawn of the AI election', 4 January 2024, https://www.prospectmagazine.co.uk/politics/64396/the-dawn-of-the-ai-election.

[213] Demos, 'Generating Democracy - AI and the coming revolution in political communications', January 2024, https://demos.co.uk/wp-content/uploads/2024/01/Generating-Democracy-Report-1.pdf.

[214] AP NORC, 'There is Bipartisan Concern About the Use of AI in the 2024 Elections', 3 November 2023, https://apnorc.org/projects/there-is-bipartisan-concern-about-the-use-of-ai-in-the-2024-elections/.

[215] The Verge, 'Republicans respond to Biden reelection announcement with AI-generated attack ad', 25 April 2023, https://www.theverge.com/2023/4/25/23697328/biden-reelection-rnc-ai-generated-attack-ad-deepfake.

In June 2023, the presidential campaign of Florida Governor Ron DeSantis used what most experts declared as AI-generated images of former President Donald Trump hugging former White House chief medical adviser Dr. Anthony Fauci in an attack ad.[216] There was no disclosure notice about use of AI and the campaign did not respond to questions.

Argentina's presidential candidates used images of themselves and each other generated by AI for their campaigns.[217] Most of the images were labelled as AI-generated or were obvious fabrications.

In Venezuela, AI was used to create fake content—a false source of information—and then to create human faces to add credibility to the message. It was echoed by the media, leaders and influencers associated with the lead political figures in what has been reported as a multi-platform influence operation.[218]

In the US, a Democratic consultant who worked for a rival presidential campaign paid a magician to use AI to create an audio clip of President Joe Biden urging New Hampshire Democrats not to vote in the state's presidential primary. This was then disseminated via robocalls. It has attracted attention from federal enforcement officials for possibly violating state voter suppression and federal telecoms laws.[219]

While the UK general election campaign is yet to begin in earnest, there is widespread concern within political parties that they will be targeted by non-party politically affiliated actors in the UK and overseas. It is also possible that central campaigns will use AI to aid existing and new political campaigning techniques, as well as to create content. In addition to Full Fact/Ipsos research mentioned in Chapter 7, which shows that a majority of the public believe generative AI and deepfakes will affect the election result, polling of MPs has also demonstrated a high level of concern.[220]

The issue has been raised by politicians in Parliament and beyond, including warnings from the shadow science and technology minister Peter Kyle about the "devastating consequences" of AI that "can further erode trust in institutions", and former justice minister Sir Robert Buckland highlighting that the "clear and present danger to

[216] The Independent, 'DeSantis campaign accused of using fake AI images of Trump hugging Fauci in ad', 8 June 2023, https://www.independent.co.uk/news/world/americas/us-politics/desantis-trump-fauci-ai-ad-b2354107.html.

[217] New York Times, 'Is Argentina the First AI Election?', 15 November 2023, https://www.nytimes.com/2023/11/15/world/americas/argentina-election-ai-milei-massa.html.

[218] Cazadores de Fake News, ¿Artificial? Sí. ¿Inteligentes? No tanto, 23 February 2023, https://www.cazadoresdefakenews.info/artificial-si-inteligentes-no-tanto/.

[219] NBC News, 'A New Orleans magician says a Democratic operative paid him to make the fake Biden robocall ', 23 February 2024, https://www.nbcnews.com/politics/2024-election/biden-robocall-new-hampshire-strategist-rcna139760.

[220] Demos, 'Generating Democracy - AI and the coming revolution in political communications', p.18, January 2024, https://demos.co.uk/wp-content/uploads/2024/01/Generating-Democracy-Report-1.pdf.

democracy presented by deepfakes and AI generated misinformation is both headed off and mitigated by direct action".[221]

> The key point is this: it is the world's biggest election year. Billions of citizens will be going to the ballot box, including here. These elections will be the first to happen since the significant advances in AI. There are legitimate concerns, anxieties and, indeed, evidence from our security services, for us to ask whether this technology will be used for fabrication, for manipulation and to affect the integrity of elections. It goes without saying that the integrity of elections matters, so that people's free choice achieves what they intend.
>
> **Chloe Smith MP, former Secretary of State for Science, Innovation and Technology**[222]

There have clearly been discussions behind the scenes but there has been a lack of leadership in this space so far from the UK's main political parties. None of them has committed to responsible use of generative AI. This is in contrast to the trend among media outlets, many of which have publicly stated how they use generative AI. Common themes in guidelines published by major publishers including the Guardian, FT, Reuters and AFP include the requirement for humans to sign off on all content, labelling uses of AI clearly to readers, and being open about intent when using AI.[223]

In the United States, several bills have been put forward, such as the The REAL Political Ads Act[224], to require disclosure of AI-generated content in political ads, and the AI Disclosure Act of 2023 and the AI Labelling Act, which both seek to have all outputs

[221] The Guardian, 'Call for action on deepfakes as fears grow among MPs over election threat', 21 January 2024, https://www.theguardian.com/politics/2024/jan/21/call-for-action-on-deepfakes-as-fears-grow-among-mps-over-election-threat.

[222] UK Parliament, House of Commons Debate: 'Political Parties, Elections and Referendums', 31 January 2024, volume 744, c. 910, https://hansard.parliament.uk/Commons/2024-01-31/debates/99BAC8DE-6003-4473-8B92-CD628AA6D859/PoliticalPartiesElectionsAndReferendums.

[223] Examples include The Guardian, ref: The Guardian, 'The Guardian's approach to generative AI', 16 June 2023, https://www.theguardian.com/help/insideguardian/2023/jun/16/the-guardians-approach-to-generative-ai; The Financial Times, ref: The Financial Times, 'Letter from the editor on generative AI and the FT', 26 May 2023, https://www.ft.com/content/18337836-7c5f-42bd-a57a-24cdbd06ec51;Reuters, ref: Thompson Reuters, *Data and AI ethics principles* (website), https://www.thomsonreuters.com/en/artificial-intelligence/ai-principles.html, (accessed 22 March 2024); and AP, ref: Associated Press, 'Standards around generative AI', 16 August 2023, https://blog.ap.org/standards-around-generative-ai.

[224] United States Congress, 'H.R.3044 - REAL Political Advertisements Act', 118th Congress (2023-2024), https://www.congress.gov/bill/118th-congress/house-bill/3044.

by generative AI labelled.[225] In February 2024, the Superior Electoral Court in Brazil set out rules for the use of artificial intelligence in political campaigning during municipal elections which take place in October.[226] This covers the use of chatbots and avatars, underpinned by the principle that such tools cannot simulate a conversation between a candidate and a real person.

The UK's next general election will happen in the coming months, so it is unlikely that regulation can be put in place in time to make a difference. In the meantime, parties must act on the concerns they have expressed, and make voluntary commitments about what they will use AI for, and how they plan to communicate when they have used AI in campaigning. Signalling commitment to standards on the use of AI may potentially deter activists from bad practices too.

Promise to use AI responsibly in election campaigns

Voters need access to accurate information in order to make informed decisions at elections. This includes information about content as well as personalisation and dissemination techniques used by parties to seek votes. With trust in politicians at its lowest level in 40 years,[227] and concern among politicians themselves, this is an opportunity to demonstrate cross-party leadership.

Ideally, AI generated content would be labelled, in line with established principles,[228] as well as published automatically to a repository. This provides a public record which can be used by journalists, fact checkers and other analysts to quickly establish whether or not content originates from parties, and how it was created. This is especially important in an information environment where content can be divorced from its original publisher and quickly travel to other places where it is loudly amplified.

Full Fact has been consulting with politicians and civil society organisations on this issue and has been working with Demos on a set of standards which we hope all political parties will be willing to commit to and implement effectively. Our political leaders must live up to the idea that with public office comes public responsibility.

[225] United States Senator Brian Schatz, 'Schatz, Kennedy Introduce Bipartisan Legislation To Provide More Transparency On AI-Generated Content', 24 October 2023, https://www.schatz.senate.gov/news/press-releases/schatz-kennedy-introduce-bipartisan-legislation-to-provide-more-transparency-on-ai-generated-content; United States Representative Ritchie Torres, 'U.S. Rep. Ritchie Torres Introduces Federal Legislation Requiring Mandatory Disclaimer for Material Generated by Artificial Intelligence', 5 June 2023, https://ritchietorres.house.gov/posts/u-s-rep-ritchie-torres-introduces-federal-legislation-requiring-mandatory-disclaimer-for-material-generated-by-artificial-intelligence.

[226] Forbes, 'Brazil Outlines Rules For AI Use During Elections', 28 February 2024, https://www.forbes.com/sites/angelicamarideoliveira/2024/02/28/brazil-outlines-rules-for-ai-use-during-elections/?sh=2bfeedc01f6a.

[227] IPSOS, 'Trust in politicians reaches its lowest score in 40 years', 14 December 2023, https://www.ipsos.com/en-uk/ipsos-trust-in-professions-veracity-index-2023.

[228] Partnership on AI, *PAI's Responsible Practices for Synthetic Media: A Framework for Collective Action* (website), https://syntheticmedia.partnershiponai.org (accessed 22 March 2024).

At a minimum, this public commitment should cover the following points:

- Parties should promise that they will not use generative AI tools to create content which is materially misleading. Anything which persuades people that something is real when it is not should be considered unacceptable. Political satire should be protected, but even in this context changes which could mislead should be highlighted as clearly as possible.
- More broadly, any content which has been altered using generative AI should be clearly labelled in such a way that is obvious to the person viewing or receiving it. Some allowance should be made for the most trivial changes, including edits which do not alter materially the implied context or content of an event. But the recent furore over the photograph[229] of the Princess of Wales and her children, which was withdrawn from use by news agencies, shows how difficult it is to draw that line.
- Parties should ensure they do not amplify synthetic content which is materially misleading, whoever has created it. They should be prepared to call out the use of such content when there is any risk that voters will be misled.
- Anyone working for or on behalf of a political party should be issued with clear public guidelines for the honest use of generative AI content in election campaigning. This process should be as transparent as possible in order to build public trust.

There is growing concern about all of these issues. In its written statement[230] to the Defending Democracy inquiry, held by parliament's Joint Committee on National Security Strategy, the Electoral Commission emphasises that "we expect anyone using AI generated campaign material to use it in a way that benefits open and transparent political debate and to label it clearly, so voters know how it has been created". Full Fact is encouraged by this, and by the Commission's call on parties and campaigners "to carry out their role influencing voters in a responsible and transparent manner".

The next general election will test whether regulation is needed

Politicians should be held to the highest standards about the use of—and transparency about the use of—AI, both as people who are asking for our trust, and as people who have power to set standards for others. Parties should sign an AI election statement as a minimum, and consider post-election what further action is needed to protect future elections. While the Electoral Commission does not regulate the content of political campaigns, there is potential for regulation of certain uses of generative AI in

[229] BBC News, 'Kate photo: Princess of Wales seen after saying she edited Mother's Day picture', 11 March 2024, https://www.bbc.co.uk/news/uk-68534359.

[230] Written evidence submitted by the Electoral Commission to the Joint Committee on National Security Strategy's Defending Democracy inquiry, 18 March 2024, https://committees.parliament.uk/writtenevidence/128808/pdf/

campaigning, particularly regarding personalisation and dissemination, but the scale of the problem is yet to be determined.

The call for political parties to use generative AI in responsible and trustworthy ways should be considered alongside wider demands for leadership on the issue of honesty in politics and election campaigning. Full Fact has called on party leaders to publicly pledge that their parties will run campaigns for the next general election honestly and transparently.[231] The rapid emergence of widely available generative AI tools has made such initiatives even more important, and emphasised how vital it is for politicians to take the lead in establishing high standards in public debate. Part two of this report sets out the changes Full Fact has been campaigning for, and recommendations for further improvements which are needed if public trust in politics and politicians is to be rebuilt.

Action for political parties

- Promise publicly to use generative AI responsibly and transparently during elections, and urge colleagues from other parties to do so too.

[231] Full Fact, 'Full Fact letter to party leaders and chairs', 2 November 2023, https://fullfact.org/media/uploads/full_fact_letter_to_party_leaders_and_chairs_on_ge24_-_november_2023.pdf.

Part 2: Trust, Politics and Government

Those in power should hold themselves to the highest standards

The general election is now less than one year away. While we do need more guardrails to protect democracy from the new risks of technology, we cannot afford to let politicians get away with deception or attempts to mislead. The introduction of a new cohort of MPs in parliament will be an opportunity to strengthen and build future systems and standards that can help to restore trust in politics.

In the second part of this report, we revisit themes from previous reports and assess where progress has been made, such as improving the parliamentary corrections system. We also analyse our fact checking corrections work to show where more needs to be done, including on the use of public information by MPs, the government and its ministers, and the widespread failure by those in public life to live up to standards of transparency and honesty.

We consider what more must happen to ensure this is not something we have to highlight in the next version of this report, including asking MPs to make good on their commitment to make it easier to correct the record, and asking ministers and government departments to be transparent about the data behind the claims they make in public.

Finally, we return to the upcoming general election, this time on the themes of honesty in campaigning and ways to embed truth and transparency in the political system once the new parliament arrives.

Chapter 9: Politicians must set the record straight

MPs need to follow through on their commitment to make it easier to correct the record (and harder to avoid doing so)

Recommendation: The Procedure Committee should finish implementing agreed changes to Parliament's corrections system without further delay, and reform the standards mechanisms for the next Parliament so that MPs who do not uphold the principle of honesty are held to account.

Put in place the new changes to the Parliamentary corrections system, and ensure all MPs use it, to normalise correcting claims

As the next general election nears, and concern about the spread of misinformation and the possible misuse of generative AI intensifies, it is even more important that MPs ensure that they use high-quality, accurate information to help inform the public when speaking in the House of Commons.

At present, Parliament continues to be a home for false, misleading and unevidenced claims. It is not possible for Full Fact to monitor and fact check the full extent of everything MPs say in the House of Commons. But in the last 12 months, MPs have allowed 14 claims to remain on the record uncorrected even after Full Fact wrote to the MPs who made them. The true number of false, misleading and unevidenced claims remaining on the record is likely to be substantially higher.

The public deserves to see honest, truthful debate, but right now the official record continues to be polluted by false, misleading or inaccurate claims from elected representatives. Following on from the last Full Fact report,[232] we have already seen out-of-date information used in legislative debate, for example by Sir Keir Starmer in December 2023 when referring to the capacity of the government's proposed scheme to send asylum seekers to Rwanda,[233] as well as factually incorrect comments made about the economy in Prime Minister's Questions.[234]

[232] Full Fact, 'Full Fact Report 2023', ch.1, March 2023, https://fullfact.org/about/policy/reports/full-fact-report-2023/report/.

[233] Full Fact, 'How many asylum seekers can the government's Rwanda scheme take?', 21 December 2023, https://fullfact.org/news/keir-starmer-rwanda-capacity.

[234] Full Fact, 'Think tank did not estimate that the average person is over £10,000 a year worse off since 2010', 13 February 2024, https://fullfact.org/economy/angela-rayner-janet-daby-disposable-incomes/.

Correcting claims needs to become the norm, but it is clear to us that this will not happen until the new Parliamentary corrections system is in place, and MPs take responsibility for ensuring that this functions effectively.

As of March 2024, all MPs, except UK Government ministers, still have to rely on making Points of Order to correct the record. This is not only an inefficient use of House time which encourages political point-scoring; it also ensures that corrections do not cross-reference to the original statements made in Hansard in a transparent way. This means that the original uncorrected claim could still appear in a search result, for example, risking the spread of bad information. An additional mechanism exists for minor corrections made by MPs,[235] but this has limited scope and, like with Points of Order, fails to correct the record in a transparent way.

In 2023, some progress was made towards tackling this issue. Full Fact and more than 50,000 signatories called for a change in the way the corrections process works.[236] This included giving all MPs equal power to correct the record in a visible way and in October 2023, MPs approved this and other changes recommended by the Procedure Committee.[237]

Six months on from the vote, change is finally on the horizon. A letter to Karen Bradley MP, Chair of the Procedure Committee, from the Editor of Hansard states that the recommendations will be implemented soon. According to the letter, they "expect the new system to go live on 15 April 2024, when the House returns after the Easter recess."[238] The following table provides an overview of the changes we expect to be made.

[235] 'Obvious mistakes' can be removed at the request of an MP, however they are notably less transparent than a ministerial correction in Hansard, where corrections use cross-referenced hyperlinks. An example of this can be seen last year, when Full Fact alerted Mary Glindon MP to a mistake that had been made by Hansard when recording a question that she asked the Prime Minister. Hansard misquoted Mary Glindon using a figure of 30%, but she had been misheard and actually said 13%. The record was subsequently corrected, but there was no public note that the correction had been made.

[236] Full Fact, 'Major victory in Parliament as Full Fact supporters ensure MPs fix broken corrections system', 25 October 2023, https://fullfact.org/blog/2023/oct/major-victory-in-parliament-as-full-fact-supporters-ensure-mps-fix-broken-corrections-system/.

[237] House of Commons Procedure Committee, 'Correcting the record', 21 June 2023, https://committees.parliament.uk/publications/40603/documents/198018/default/.

[238] House of Commons Procedure Committee, letter received regarding 'Correcting the record (Fourth Report of Session 2022-23)', 29 February 2024, https://committees.parliament.uk/publications/43776/documents/217297/default/.

Expected changes to corrections process in Parliament

Procedure Committee recommendations	Changes
Cross-referenced hyperlinks provided in the Official Report should be improved…through replacing the existing code with wording clearly stating that the link directs the reader to a correction.	Ascription links directing readers to a correction will include the word “Correction” following the column reference.
Cross-referenced hyperlinks currently used in the ministerial corrections system should also be added to corrections made through points of order and other oral contributions.	Where a member states on the Floor of the House that they wish to correct a remark they have made and refers to the Official Report by date and column reference, an ascription link will be included both in that statement and alongside the original words spoken, where it will include the word “Correction”. The members’ Guide to Procedure will include guidance on this point and will also be updated to reflect the other changes outlined in this letter
Corrections should be easier to access…through the creation of a central corrections page [including] all corrections that have been made, including through written ministerial statements and points of order, in chronological order, with the topic and link included.	A corrections page will be published on parliament.uk and updated weekly with any corrections made by ministers or other members, including through written ministerial statements and points of order, in chronological order, with the topic and link included.
The most effective option to improve opportunities for backbenchers to correct the record is to incorporate them fully into the existing ministerial corrections system…They should also be required to adhere to the same standards as set in the ministerial corrections system.	The Hansard daily part and web pages will include a section entitled “Written corrections” subdivided into “Ministerial corrections” and “Other corrections”. The latter will include corrections received from members who are not ministers. The column reference used to refer to written corrections will change to WC to reflect their broader compass.

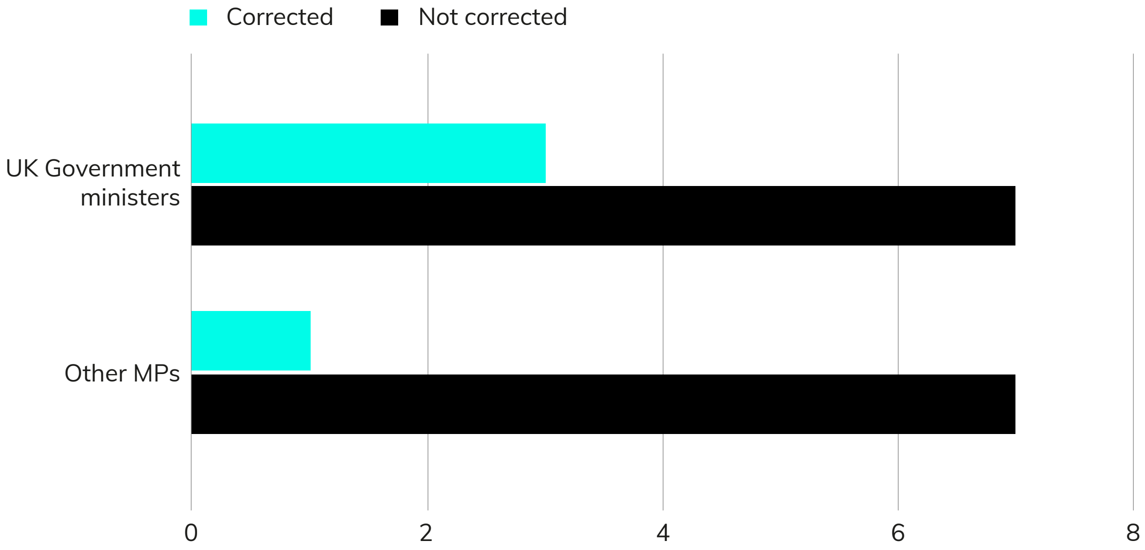
We are pleased to see that a new edition of the MPs’ Guide to Procedure[239] has been published in anticipation of the new corrections system. However, it is important that the implementation of this system and the other changes outlined above are just the first and not the final steps to stop and reduce the spread of bad information from within Westminster. For example, corrections still need to be clearly flagged in audio and video footage shared by Parliament.

When the ministerial-style corrections system is extended to all MPs—as already approved—corrections will happen more transparently and, we hope, more consistently. Without the restriction of having to make a Point of Order, MPs should welcome the implementation of the new system and set an example by using it, when asked, to ensure that information in the official parliamentary record is trustworthy.

[239] House of Commons, ‘MPs’ Guide to Procedure’, January 2024, https://guidetoprocedure.parliament.uk/dist/mps-guide-to-procedure.pdf.

Evidence from Full Fact's correction requests in the last 12 months suggests that, while the ministerial correction system is far from a perfect mechanism, the ease and transparency with which a minister can correct the record means that more corrections are actually made. The chart below shows that when we have asked for the record to be corrected in Hansard, ministers have done so more reliably than other MPs, although there still is considerable scope for improvement, and there may be factors besides the corrections system in play.

Outcomes after Full Fact requested a correction to the record in Hansard during the last 12 months to March



Full Fact hopes that the number of corrections will increase significantly when the system is available to all MPs.

The Committee on Standards must hold MPs to account when they refuse to correct persistent or egregious misleading claims

The proposals outlined above are entirely voluntary. The House of Commons does not currently have a system to tackle consistent or egregious failures by ministers and other MPs to correct their mistakes. This must be remedied if Parliament has any desire to act on the public's call for politicians to face consequences for dishonesty.[240]

[240] "74% of contributors say democracy in the UK could be improved if MPs were 'thrown out' of Parliament for lying or faced some form of consequences for their actions (Renwick et al., 2023b)", ref: UK Governance Project Commission, 'Governance Project', p.57, 1 February 2024, https://www.ukgovernanceproject.co.uk/wp-content/uploads/2024/02/Governance-Project-Final-Report-31.1.24.pdf.

MPs are obliged to abide by a Code of Conduct,[241] inspired and informed by the Seven Principles of Public Life.[242] MPs should "robustly support the principles", "be truthful" and "submit themselves to the scrutiny necessary." But in too many cases, MPs do not.

There is also a Ministerial Code[243] which directs ministers to observe the Seven Principles of Public Life. The third principle states that it is "of paramount importance that ministers give accurate and truthful information to Parliament, correcting any inadvertent error at the earliest opportunity. Ministers who knowingly mislead Parliament will be expected to offer their resignation to the Prime Minister".

But Chart 1 shows that in the 12 months to March 2024, the majority of UK Government ministers and MPs failed to correct a false or misleading claim after being asked to do so by Full Fact.

Under current procedures, even a Prime Minister—to whom the Ministerial Code applies—can ignore requests to correct the record. This has happened even when a claim has prompted action by Full Fact, the Office for Statistics Regulation, the UK Statistics Authority, the Liaison Committee, and over 18,000 members of the public.[244] Unfortunately, there are no meaningful repercussions for a Prime Minister, other ministers or MPs who fail to correct a false or misleading claim soon after being informed about it.

Full Fact provided evidence to the Committee on Standards in Public Life's Inquiry into the Standards Landscape in 2023. We made the case for four key recommendations which would help to ensure that MPs and ministers are held to account effectively for refusing to correct consistent or egregious misleading claims. These recommendations should be implemented before the next parliament sits:

- Introduce a new streamlined process to deal with MPs who make consistent or egregious misleading claims and refuse to correct them, and ensure they are held to account effectively
- Work closely with the Prime Minister to consider new mechanisms to ensure that the Ministerial Code of Conduct is more stringently enforced, including when ministers fail to correct their mistakes in Parliament
- Commit to a future inquiry on how to deal with MPs who make false or misleading claims outside of Parliament, and on the role the Parliamentary Commissioner should have within this

[241] House of Commons, 'The Code of Conduct', 12 December 2022, https://publications.parliament.uk/pa/cm5803/cmcode/1083/1083.pdf.

[242] Committee on Standards in Public Life, 'Guidance: The Seven Principles of Public Life', 31 May 1995, https://www.gov.uk/government/publications/the-7-principles-of-public-life/the-7-principles-of-public-life--2.

[243] UK Government Cabinet Office, 'Ministerial Code', December 2022, https://assets.publishing.service.gov.uk/media/63a4628bd3bf7f37654767f2/Ministerial_Code.pdf.

[244] Full Fact, 'Here to lead not mislead', 20 April 2022, https://fullfact.org/blog/2022/apr/here-to-lead-not-mislead/.

- Develop, under the House Service, in-depth training on standards that should be delivered to all MPs within six months of a general election, and for new MPs within six months of their election

Further detail about how and why these recommendations would work is available in the evidence we provided.[245] For example, we explained how the proposed in-depth training would help to address the lack of understanding some MPs have about the rules on correcting mistakes in Parliament, and on upholding the standard of Honesty, as set out in the Seven Principles of Public Life. The training should also include guidance on sharing and presenting information accurately, and how to pursue a correction within the present system (albeit in a standards landscape which urgently needs to be improved).

MPs must uphold the principle of honesty beyond the Commons, and correct false or misleading claims made in the media or online

Creating a better process for correcting the record in the House of Commons is just one important part of what will be required to establish greater honesty and accuracy in British politics, and to rebuild public trust.

It is also vital to deliver improvements outside Parliament—be it in a speech, a newspaper article, a television or radio interview, or other online communication—whenever MPs are acting in their capacity as public representatives. It is worth emphasising that when public attention is focused more intensely on politics, as it will be during the upcoming general election, there is a heightened risk that falsehoods and inaccuracies by MPs will spread more widely via communication outside the Commons than within.

When an MP is informed by Full Fact that they have made a false or misleading claim outside the Commons, we routinely ask them to correct it and make efforts to ensure that anyone who might have heard the claim is aware of the correction. We ask for the broadcaster of the claim to be notified, and also for the correction to be communicated via the claimant's own channels, such as their social media accounts.

For example, in January 2024 during an interview on GB News, Conservative MP Jonathan Gullis repeated a misleading claim about Labour's immigration plans.[246] A clip from the interview was then shared in a paid-for Facebook advert. We asked Mr Gullis to issue a correction on Facebook and also make GB News aware, but he has yet to respond to our request. Similarly, in October 2023 we asked the shadow chancellor,

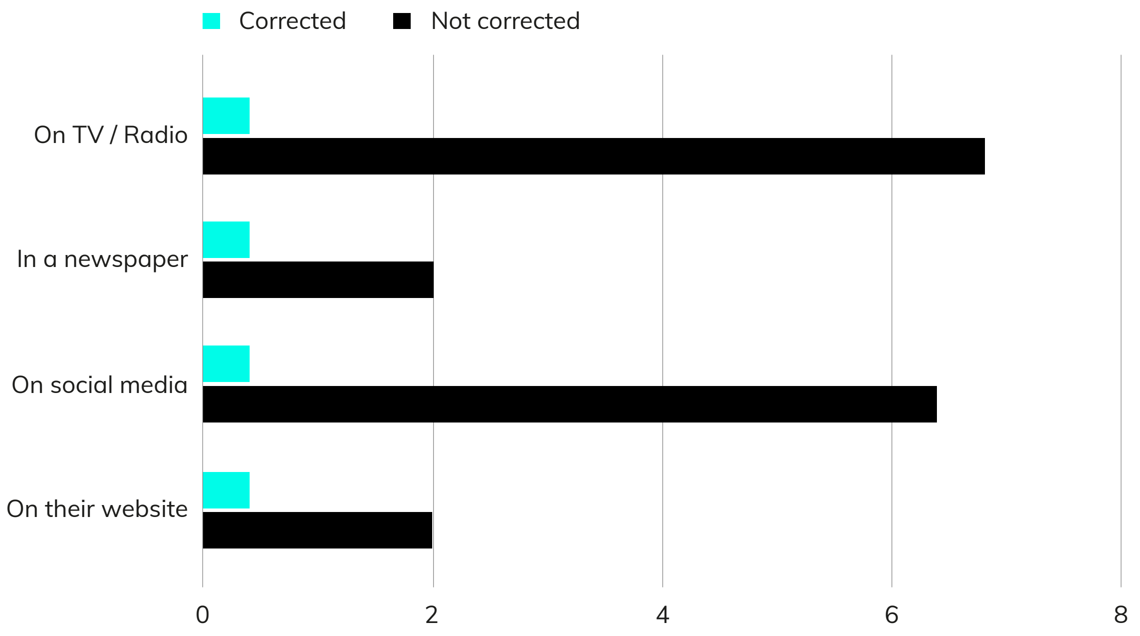
[245] House of Commons Committee on Standards Inquiry on the House of Commons Standards Landscape, 'Written evidence from Full Fact', 2023, https://committees.parliament.uk/writtenevidence/125122/pdf/.

[246] Full Fact, 'Local Conservative Facebook ad featured misleading "100,000 migrants" claim', 19 January 2024, https://fullfact.org/news/2023-24-parliamentary-session/liveblog-jonathan-gullis-political-ad/.

Rachel Reeves, to correct a claim about house building levels that she made in an interview with the Sunday Times, but she did not respond.[247]

As the chart below shows, Mr Gullis and Ms Reeves are far from alone. Nowhere near enough claims communicated outside the Commons by MPs are currently being corrected.

Outcomes after Full Fact requested a correction to a claim made by an MP during last 12 months to March 2024



Parties must take responsibility for what their MPs or Parliamentary candidates say, wherever they say it. With this in mind, we have invited parties with representatives in the House of Commons to publicly pledge to run their campaigns for the next general election honestly and transparently, making sure that the claims they and their candidates make are truthful.

If the main political parties make this pledge (see Chapter 11), and continue to fulfil it beyond the election, we hope to report a significant reduction in the number of uncorrected claims by MPs on their social media accounts, as well as their websites and other media.

[247] Full Fact, 'House building figures not at lowest level since the Second World War', 10 October 2023, https://fullfact.org/news/rachel-reeves-housebuilding-war.

Media outlets and their regulators must do more to ensure that MPs quickly correct false and misleading claims

While the responsibility remains with elected officials to ensure the information they are sharing is honest and accurate, broadcasters and newspapers can also do more to help avoid the spread of bad information.

Many media outlets issue prompt corrections when contacted by us, which is consistent with principles of accuracy set out in relevant regulatory and editorial codes such as Ofcom's Broadcasting Code,[248] the Independent Press Standards Organisation's (IPSO) Editors' Code of Practice[249] and the BBC's Editorial Guidelines.[250] But this practice is not yet the norm.

If all media outlets can ensure that (i) MPs are held to account at the earliest opportunity for claims they make, and (ii) correction requests are swiftly processed and promoted—preferably referencing wherever the claim was accessible, such as in the description of a video/audio recording—this will help facilitate a culture of transparency and echo the changes that could soon be seen in Parliament. The media plays a crucial role in seeking due accuracy by holding MPs to account, so the scrutiny and corrections processes they offer should at least be on a par with what Parliament should provide.

When goodwill and internal correction processes do not go far enough, regulators can play an impactful role in sanctioning bad information, but there is scope to go further. For example, IPSO, the independent regulator of most of the UK's newspapers and magazines, has the power to secure corrections through the Editors' Code of Practice, but due to the time taken to assess such proceedings, false and misleading claims often remain publicly available for long periods before they are corrected. It was reported, for example, that IPSO took more than four months to uphold a complaint about an inaccurate article that was used by a major political party during the 2019 general election campaign.[251]

We encourage regulators to take the lead in identifying ways to significantly reduce the time required for decisions to be reached on complaints about claims made in the media by political representatives, and to ensure they are acted upon swiftly.

[248] Ofcom, Broadcasting Code, 'Section five: Due impartiality and due accuracy', 5 January 2021, https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-codes/broadcast-code/section-five-due-impartiality-accuracy.

[249] IPSO, 'Editors' Code of Practice', January 2021, https://www.ipso.co.uk/editors-code-of-practice/.

[250] BBC, *Editorial Guidelines, Section 3: Accuracy - Introduction* (website), https://www.bbc.co.uk/editorialguidelines/guidelines/accuracy/ (accessed 22 March 2024).

[251] The Guardian, 'Mail on Sunday made false claims about Labour's tax plans', 9 December 2019, https://www.theguardian.com/media/2019/dec/09/ipso-rebukes-mail-on-sunday-over-labour-movers-tax-claim.

At a time when trust in parties, the government and parliament is so low,[252] it is essential that steps are taken to rebuild it. Full Fact wants to support that process, and we will highlight positive moves when they are made. But there is a risk that honesty and accuracy in British politics will decline further unless Parliament, MPs, the media and regulators all work proactively with us and other concerned organisations to achieve progress.

Action for Parliament

- The Procedure Committee must finish implementing the agreed new changes to Parliament's corrections system, and ensure that corrections are communicated transparently on Parliament's audio and video channels.
- Introduce a new streamlined process to effectively hold MPs to account when they make consistent or egregious misleading claims and refuse to correct them.
- The Committee on Standards in Public Life should work closely with the Prime Minister to consider new mechanisms to ensure that the Ministerial Code is more stringently enforced, including when ministers fail to correct their mistakes in Parliament.
- The Committee on Standards in Public Life should hold a future inquiry on how to deal with MPs who make false or misleading claims outside Parliament, and the role of the Parliamentary Commissioner in this.
- The House Service should develop in-depth training on standards to be delivered to all MPs within six months of a general election, and for new MPs within six months of their election. This should include upholding the principle of Honesty as described in the Seven Principles of Public Life, and mechanisms for how to correct mistakes and pursue a correction from others.

Action for MPs

- MPs must fully adhere to their Code of Conduct and the Seven Principles of Public Life, as well as the Ministerial Code where applicable.
- MPs should call upon their leaders to publicly pledge to run their campaigns for the next general election honestly and transparently, and honour this pledge themselves.
- When an MP is reliably informed that they have made a false or misleading claim, they should correct it and make efforts to ensure that anyone who might have heard the claim is aware of the correction.

[252] Office for National Statistics, 'Trust in government, UK: 2023', 1 March 2024, https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/bulletins/trustingovernmentuk/2023.

Action for IPSO and Ofcom

- IPSO and Ofcom should find new ways to significantly reduce the time required for decisions to be reached on complaints about claims made in the media by MPs.

Action for the media

- All broadcasters and newspapers should adhere to the standards set out in the Broadcasting Code253 and Editors' Code of Practice,254 ensuring due accuracy by holding MPs to account at the first opportunity for claims they make.
- All broadcasters and newspapers should ensure clear, responsive mechanisms for processing requests to correct claims made by politicians and parties.

[253] Ofcom, 'The Ofcom Broadcasting Code', 31 December 2020, https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-codes/broadcast-code.

[254] IPSO, 'Editors' Code of Practice', January 2021, https://www.ipso.co.uk/editors-code-of-practice/.

Chapter 10: Government must provide evidence for every claim that it makes

Ministers and government departments must be fully transparent about the data behind any claim that they make

Recommendation: Ministers and government departments must provide evidence for what they say, and use public data in line with the Code of Practice for Statistics. This must be embedded in the Ministerial Code, and Parliament must hold ministers to account when they fail to live up to these standards.

Government must be transparent in its use of data and provide it in an easy-to-understand format

We are repeating, word-for-word, the recommendation we made in the 2023 Full Fact report because nothing has changed.[255] Last year we highlighted that government departments, and government ministers in particular, are sometimes too quick to throw around numbers to support their claims and too slow to publish the important supporting or contextual data behind them.

Full Fact's work continues to expose inadequate government use of data, as we set out below, and this has damaging consequences. When the government fails to provide evidence to back up its claims, it harms public confidence in democracy. It should be a grave cause for concern that nearly two thirds (62%) of the UK population feels that the government is actively misleading them by making claims it knows to be false.[256]

To restore public confidence in our democracy it is essential that the government is held to the highest standards of honesty and accuracy.

[255] Full Fact, 'Full Fact Report 2023', ch.1, March 2023, https://fullfact.org/about/policy/reports/full-fact-report-2023/report.

[256] Edelman, 'We will never realise the promises of the future without trusted information', 23 January 2024, https://www.edelman.co.uk/research/future-without-trusted-information.

What we have seen in the last 12 months

Full Fact's work in 2023 and early 2024 demonstrates poor government use of data to back up the claims that it makes. These fall into three broad categories:

- Claims based on unpublished data
- Claims based on non-existent data
- Claims based on selective use of data

These behaviours contravene the Code of Practice for Statistics.[257] The Code of Practice—adherence to which is currently voluntary— is overseen by the Office for Statistics Regulation (OSR) and the independent regulator of the UK Statistics Authority. The OSR provides "independent regulation of all official statistics produced in the UK… to enhance public confidence in the trustworthiness, quality and value of statistics produced by government."[258]

A report published by Full Fact in June 2023[259] highlighted the repeated misleading use of data by government ministers, with claims being made by the Home Office based on unpublished operational data a source of particular concern. For example, we fact checked then immigration minister Robert Jenrick's unevidenced claim in Parliament in November 2022 about the percentage of adult men arriving at Western JetFoil asylum processing centre that are claiming to be under 18.[260] The Home Office told us that this claim was based on provisional operational data, which had not been made public. Despite our attempts, we were not able to obtain this data, and were not able to establish if what Mr Jenrick said was accurate.

Similarly, in March 2023, there was no published data to support the Prime Minister's claim that there were 6,000 fewer people in the caseload of the asylum backlog.[261] We wrote to Mr Sunak to ask for the source of his claim, but did not receive a response. The Home Office subsequently started publishing ad hoc data which suggested he may have been referring to the "legacy backlog" which is a subsection of the overall backlog of cases. Full Fact has argued repeatedly that the government has a responsibility to ensure that the information, statistics and analysis it publishes is presented transparently, and that trust is severely undermined when official information is found to be unevidenced, lacking the full context or misleading.

[257] Office for Statistics Regulation & UK Statistics Authority, 'Code of Practice for Statistics', 5 May 2022, https://code.statisticsauthority.gov.uk/.

[258] Office for Statistics Regulation, 'Annual Report 2022/23', 13 July 2023, https://osr.statisticsauthority.gov.uk/publication/office-for-statistics-regulation-annual-report-2022-23/.

[259] Full Fact, 'Government Statistics: misrepresentation and data gaps', June 2023,https://fullfact.org/media/uploads/government_statistics_-_misrepresentation_and_data_gaps.pdf.

[260] Full Fact, 'No published data to support minister's claim about migrants saying they're under 18', 22 November 2022, https://fullfact.org/immigration/robert-jenrick-fifth-male-migrants-under-18/.

[261] Full Fact, 'No evidence to support Rishi Sunak's asylum backlog claim', 10 March 2023, https://fullfact.org/immigration/rishi-sunak-asylum-backlog/.

We also see examples of the use of unpublished data in a health context. In October 2023, for example, then health secretary Steve Barclay made the claim that the number of NHS patients in Wales travelling over the border to England in order to receive treatment had increased by 40% in the past two years.[262] However, this assertion appeared to be based on unpublished data, and when we asked for the source, Mr Barclay did not respond. NHS England does collect such data but does not routinely publish it. We did however discover that a similar statistic to that used by Mr Barclay was reported in the media based on "internal NHS figures" for 2020/21 and 2022/23.[263]

This behaviour from Mr Barclay is a clear contravention of the Code of Practice for Statistics principle T3.8, which states:

> …ministerial statements referring to regular or ad hoc official statistics should be issued separately from, and contain a prominent link to, the source statistics. The statements should meet basic professional standards of statistical presentation, including accuracy, clarity and impartiality.[264]

The use of unpublished information, which is impossible to verify independently, can erode public trust in authorities. It should not be left to fact checkers and journalists to uncover the evidence underpinning government claims. Such absence of published evidence prevents public scrutiny, which should be a basic requirement of any democratically elected government. Currently, the Ministerial Code states that ministers only need to "be mindful of" the Code of Practice for Statistics. We believe that ministers should be required to adhere to the Code of Practice.

It is a cause of concern when unpublished data is cited, but it is arguably worse when the government makes unevidenced claims where the underlying data may not even exist. For example, in July 2023, the Prime Minister claimed that A&E waiting times in England were the "best in two years".[265]

[262] Full Fact, 'Are 40% more Welsh patients 'escaping' to England for treatment?', 12 October 2023, https://fullfact.org/health/wales-nhs-england-treatment-steve-barclay/.

[263] MailOnline, 'Patients try to "escape" Labour's Welsh NHS as the number seeking care in English hospitals to avoid longer waits rises by almost 40% in two years' 13 August 2023, https://www.dailymail.co.uk/news/article-12402927/Patients-escape-Welsh-NHS-England-hospitals.html.

[264] Code of Practice for Statistics, T3.8 p.20, https://code.statisticsauthority.gov.uk/.

[265] UK Parliament, Prime Minister's Questions, HC Volume 736 Column 900, 19 July 2023, https://hansard.parliament.uk/Commons/2023-07-19/debates/C7194093-D8DD-4DA3-9F5A-61B025CA09C2/Engagements.

We asked Downing Street where they got the evidence for this claim. We did not receive a response. When we consulted health experts, they found the claim inconsistent with published data which, while it did indicate some performance improvements, did not suggest waiting times were the best they'd been for 24 months. Instead, the 12-month averages for three key metrics on waiting times had all worsened since June 2021.[266]

Ultimately, in the absence of any evidence provided to the contrary, we concluded that data probably did not exist to support the claim that A&E waiting times in England were the "best in two years".[267] It is unacceptable for the government to behave in this manner and is a clear breach of the Nolan principles of Honesty and Openness.

We've seen how the government makes claims based on unpublished data, and where the data is non-existent. In other examples, the data can seem to be accurate at first glance, but still be misleading, because ministers may make claims that stand up to initial scrutiny of the data underpinning them, yet are missing essential context. Our work in the past year has demonstrated that context is vital. It appears that the government is often prepared to leave out key data in order to make the best possible case on certain policies. We've seen versions of this in a number of areas. For example, in November last year, the Prime Minister claimed that the UK was doubling aid for Palestinian civilians.[268] It is true that recently announced aid for the Occupied Palestinian Territories (OPTs) doubled the total amount in aid commitments for the OPTs this year. However this comes after aid commitments to the OPTs—and in general—have reduced substantially in recent years.

The same tactic has been used when talking about house building. In July 2023, Rishi Sunak claimed that the Conservative Party was responsible for "record levels" of house building during its period in government.[269] The claim used a metric known as "net additional dwellings"—but this metric also includes conversions of office buildings and the subdivision of houses into flats, as well as actual house building. A different dataset, "indicators of new supply", counts the number of completed new dwellings, and is perhaps more in line with what the public might consider to be "new" houses. This metric told a different story, and didn't support the claim that house building was at or close to record levels.[270]

The public should be able to take statements from the government at face value without having to search out the additional context.

[266] A&E four-hour performance in England, Patients waiting to be admitted to hospital from A&E, and Patients spending more than 12 hours in A&E in England.

[267] Full Fact, 'Data doesn't seem to back up PM's claim that A&E waits are "the best in two years"', 9 August 2023, https://fullfact.org/health/accident-and-emergency-rishi-sunak-nhs-waiting-times/.

[268] Full Fact, 'How has UK aid spending for Palestinians changed in recent years?' 13 November 2023, https://fullfact.org/news/palestinian-aid-spending/.

[269] Full Fact, 'Conservative and Labour claims on house building fact checked', 12 July 2023, https://fullfact.org/economy/house-building-levels-PMQs/.

[270] Ibid.

Full Fact is far from alone in having serious concerns about government misuse of statistics. In May 2023, Ed Humpherson, the Director General for Regulation at the Office for Statistics Regulation (OSR) wrote to Full Fact to say that he shared our disappointment that claims were continuing to be made by ministers that could not be verified from the Home Office's published statistics.[271] He informed us that the OSR had been engaging with the Home Office both publicly and privately for some time, and its response had been constructive. More recently, the UK Statistics Authority, in its response to a complaint about the government's claims around the asylum backlog, has also raised the alarm about ministers not taking on board the advice of lead statisticians when preparing communications.[272]

To fix this, there needs to be proper checks and balances in place. To begin with, mandatory adherence to the Code of Practice for Statistics, and embedding it in the Ministerial Code, would provide better oversight of government use of data, and provide a method of enforcement on those occasions when things go wrong. In the next chapter we explain why the Ministerial Code needs to be independently overseen, and placed on a statutory footing. To further help transparency, all government departments' annual reports should reference any concerns raised by the OSR related to their use of statistics. Meanwhile, Parliament needs to exercise its own powers of scrutiny more effectively, through the Select Committee system. Departmental committees must ensure that when they regularly question the relevant ministers and civil servants, any recorded misuse of statistics is on the agenda.

Action for the government

- Strengthen the Ministerial Code to make it clear that ministers should adhere to the principles of the Code of Practice for Statistics for any data they use to back up statements they make.
- Comply with the OSR recommendation that any quoted data used to back up a government claim is published in an "equally accessible" format, and furthermore, make it a statutory requirement.[273]
- Each government department's annual report should highlight any concerns raised publicly by the OSR and set out the department's response.

[271] Office for Statistics Regulation, 'Ed Humpherson to Will Moy: Home Office Transparency', 23 May 2023, https://osr.statisticsauthority.gov.uk/correspondence/ed-humpherson-to-will-moy-home-office-transparency/.

[272] UK Statistics Authority, 'Response from Sir Robert Chote to Alistair Carmichael MP – Asylum backlog figures', 18 January 2024, https://uksa.statisticsauthority.gov.uk/correspondence/response-from-sir-robert-chote-to-alistair-carmichael-mp-asylum-backlog-figures/.

[273] Office for Statistics Regulation, 'Statement on data transparency and the role of Heads of Profession for Statistics', 13 July 2021, https://osr.statisticsauthority.gov.uk/news/osr-statement-on-data-transparency-and-the-role-of-heads-of-profession-for-statistics/.

Action for Parliament

- Parliament and select committees should take a more active role in scrutinising and holding ministers and government departments to account about the way they evidence their claims.

Chapter 11: Strengthen the culture and system to create more trust in politics

The public deserves an honest election, and a future government and parliament that will help restore trust in our politics

Recommendations: All political parties must commit to honest campaigning during the next election. The next government should legislate to end deceptive campaign practices, introduce independent regulation of political advertising and put the Ministerial Code on a statutory footing.

Trust in politicians is worryingly low

Change cannot come soon enough. Public opinion about politicians' capacity for being honest is at a record low: only nine per cent of British adults trust politicians to tell the truth.[274] According to new research conducted by Ipsos UK for Full Fact, a majority of the UK public (71%) are concerned that voters will be misled by false or misleading claims in the upcoming election campaign. The same proportion of the population is supportive of political parties adopting a set of standards for honesty and transparency in manifestos: 71% agree or strongly agree with such an idea (32% strongly).[275]

In the previous two chapters, we have looked at why politicians must correct the record when they make mistakes, and why the government must be as transparent as possible with the data it uses to back up its claims. With an election and a new Parliament on the horizon, this chapter explores how the conduct of the election campaign and the action taken afterwards can produce positive changes for the future, to help rebuild trust in the system.

[274] Ipsos, 'Trust in politicians reaches its lowest score in 40 years', 14 December 2023, https://www.ipsos.com/en-uk/ipsos-trust-in-professions-veracity-index-2023.

[275] Ipsos, 'Full Fact UK Public Attitudes Research', April 2024, http://fullfact.org/audience-research-2023.

Full Fact has asked parties to help restore public trust in political campaigning

There will be a general election by January 2025 at the latest and all eyes will be on politicians and their behaviour. In Part 1 we looked at the likely impact of AI and misinformation on the election and party campaigning. In addition, we are likely to witness a more traditional array of tactics as parties compete to gain our trust and our votes, from campaign leaflets disguised as local newspapers, to manifestos which contain uncosted or unrealistic promises.[276]

Full Fact is campaigning for party leaders to promise to do better this time around. We have asked that parties pledge to:

- Make sure that the claims made by the party, its leader and its candidates are truthful
- Set out the party's manifesto in ways that allow meaningful scrutiny of its pledges
- Ensure the party's advertising is honest and truthful, and commit to have the party's political advertising independently regulated in the future
- Not use deceptive campaigning tactics to gain votes, and commit to new rules for honest party campaigning practices

At the time of writing in March 2024, the Green Party, the Alliance Party, the SDLP and Plaid Cymru had agreed to this pledge.[277] Full Fact is in conversation with other parties, but it remains to be seen whether any of them will also decide to sign.

Parties should agree to independent regulation of political advertising

When reviewing a selection of political advertising from the 2019 general election campaign, the campaign group Reform Political Advertising found multiple instances of misleading or exaggerated claims.[278] Furthermore, their report on the 2022 local elections, for which Full Fact formed part of their Election Advertising Review Panel, observed an “alarming amount of grossly misleading election advertising from all main parties”.[279]

[276] Full Fact, ‘It's time political parties tidied up their election campaigns’,10 November 2023, https://fullfact.org/blog/2023/nov/letter-to-political-parties/.

[277] Full Fact, ‘Major parties commit to standards for honest politics for next election’, 17 January 2024, https://fullfact.org/blog/2024/jan/major-parties-commit-to-standards-for-honest-politics-for-next-election/.

[278] Reform Politicial Advertising, ‘Illegal Indecent Dishonest and Untruthful’, December 2019, https://reformpoliticaladvertising.org/wp-content/uploads/2019/12/Illegal-Indecent-Dishonest-and-Untruthful-The-Coalition-for-Reform-in-Political-Advertising.pdf.

[279] Reform Political Advertising, ‘The Cost of Lying Crisis’, 5 May 2022, https://reformpoliticaladvertising.org/wp-content/uploads/2022/05/Reform-Political-Advertising.-COST-OF-LYING-CRISIS-1.pdf.

The Advertising Standards Authority (ASA) ceased to oversee most political advertising in 1999.[280] Since then, political advertising “where the principal function is to influence voters in a local, regional, national or international election or referendum” has been exempt from the UK Code of Non-broadcast Advertising and Direct & Promotional Marketing (CAP Code).[281] This has allowed significant misinformation in advertising to go unchecked, for example on Facebook, as we found during the 2019 election.[282]

Political ads should be subject to the same standards of factual accuracy and evidence to which other advertising must adhere. Without the ASA having oversight, parties must commit to the creation of a new independent regulator specifically focused on political ads.

Full Fact is not alone in making this call, nor is this unprecedented. A regulatory system has been introduced in New Zealand,[283] and others such as Reform Political Advertising,[284] and the Neill Committee as long ago as 1998,[285] have made similar suggestions in the UK.

Beyond advertising, we need improvements to rules on campaign materials to prevent deceptive behaviour, such as disguising the provenance of electoral material, or presenting it as something separate and independent like a local newspaper.[286] The size of an imprint, which makes it clear that the material has been produced by a political party, is particularly important. The next government should consider introducing targeted legislation to stamp out obvious cases of deception.

After the election our new government and parliament must embed truth and transparency in the political system

Once the election is over, there will be an opportunity for the next government and Parliament to make a fresh start on transparency.

[280] Advertising Standards Authority, ‘Why we don’t regulate political ads’, 26 April 2023, https://www.asa.org.uk/news/why-we-don-t-regulate-political-ads.html.

[281] Advertising Standards Authority, ‘Think you know what the CAP Code applies to?’, 20 June 2019, https://www.asa.org.uk/news/think-you-know-what-the-cap-code-applies-to.html.

[282] Full Fact, ‘The facts behind Labour and Conservative Facebook ads in this election’, 11 December 2019, https://fullfact.org/election-2019/ads/.

[283] Campaign Live, ‘No more ‘freedom to lie’: follow New Zealand’s example to reform UK political advertising’, 19 July 2021, https://www.campaignlive.co.uk/article/no-freedom-lie-follow-new-zealands-example-reform-uk-political-advertising/1722381.

[284] Reform Political Advertising (website), https://reformpoliticaladvertising.org/ (accessed 22 March 2024).

[285] House of Commons Library, ‘Who regulates political advertising?’, 4 November 2019, https://commonslibrary.parliament.uk/who-regulates-political-advertising/.

[286] Full Fact, ‘Deceptive Campaign Practices – FAQs’, November 2023, https://fullfact.org/get-involved/petitions/end-deceptive-campaigning/faqs/.

Our political system is based on convention, and a gradual evolution of parliamentary and governmental practices and culture. John Major's government made a code of practice for ministers public for the first time in 1992,[287] and established the Committee on Standards in Public Life in 1994. This Committee then devised the Seven Principles of Public Life in 1995, sometimes known as the Nolan Principles, after the first chair of the Committee[288]. Three of these principles are of particular interest to Full Fact: Honesty, Accountability, and Openness.

The official Ministerial Code[289] (renamed under the Blair Government in 1997) has no legal basis but has become an accepted set of standards for ministerial behaviour. When a breach is alleged to have taken place, it is at the Prime Minister's discretion to decide whether and how it is investigated.

As well as promoting adherence to the Nolan Principles, the terms of the Code create an expectation that ministers will resign if they knowingly mislead Parliament. The Code contains the following clause:

> “It is of paramount importance that Ministers give accurate and truthful information to Parliament, correcting any inadvertent error at the earliest opportunity.[290]

However, expectation is not enough. As set out in Chapter 9, the majority of government ministers and MPs who have been asked by Full Fact to correct the record have failed to do so. As well as clearly being a breach of the clause quoted above, a failure to correct the record also demonstrates a failure to uphold the Nolan principles of Honesty and Openness. Without any independent arbitration, or any compulsion for ministers to adhere to the Code, we are left with, in effect, governments and political parties that are able to mark their own homework.

Independent oversight is clearly necessary. We explained in Chapter 9 that there have been seven occasions since 2022 where the current Prime Minister has been asked to correct the record following Full Fact identifying errors he had made. To date, he has managed to do so only once.[291] If the Prime Minister does not set the right example by correcting the record—and acts in breach of their own Ministerial Code—they should certainly not be the one overseeing it.

[287] Institute for Government, 'Ministerial code', 26 April 2019, https://www.instituteforgovernment.org.uk/explainer/ministerial-code.

[288] House of Commons Library, 'Seven Principles of Public Life', 24 August 2022, https://commonslibrary.parliament.uk/research-briefings/cdp-2022-0156/.

[289] UK Government, 'Ministerial Code', 22 December 2022, https://www.gov.uk/government/publications/ministerial-code/ministerial-code.

[290] UK Government, 'Ministerial Code', 22 December 2022, https://www.gov.uk/government/publications/ministerial-code/ministerial-code.

[291] Full Fact (website), 'Rishi Sunak MP', https://fullfact.org/can-i-trust/1077/rishi-sunak(accessed 22 March 2024).

This cannot continue. The next government should restore respect for Parliament and make clear that there will be zero tolerance for ministers making misleading statements or failing to correct their mistakes. The Ministerial Code should be placed on a statutory footing and overseen independently.

While the Ministerial Code only applies to government ministers, the Nolan Principles apply to all Members of Parliament through the House of Commons Code of Conduct for MPs.[292] Both the Ministerial Code and the Code of Conduct for MPs provide a useful set of standards, but they are often not adhered to—and when they are breached, action is not always taken.

So there is a twofold problem: poor behaviour with regards to honesty and transparency, and lack of enforcement of standards designed to prevent such behaviour.

For the former, there must be higher standards demanded as soon as someone first becomes an MP. All MPs should make a declaration to be honest and tell the truth alongside swearing an oath to the King when they enter Parliament. A similar process already happens in the House of Lords, whereby Peers agree to abide by the House of Lords Code of Conduct immediately after making their oath to the King.[293] Such a change would not require any new legislation, but could for example be introduced through a recommendation from the Commons Procedure Committee, and with the blessing of the Speaker it could then be put to a vote in the House.

Furthermore, new MPs should undergo training on standards within six months of their election, and any such training should include guidance on sharing and presenting information accurately, and on how to pursue a correction from a fellow MP. This was discussed during a recent oral evidence session of the House of Commons Committee on Standards. During that session, Leader of the House Penny Mordaunt MP accepted that there was a need for better training due to the varied nature of the standards with which MPs must comply:

> ...there are an enormous number of standards bodies... A quick count brings up 13 different organisations, but most Members of Parliament are not really sighted on those bodies. It is only if they encounter them in some particular capacity that they know about them. The rules are very opaque. There is not a great deal of training, or a one-stop shop where people can go to look at those things.[294]

[292] House of Commons, 'The Code of Conduct', 12 December 2022, https://publications.parliament.uk/pa/cm5803/cmcode/1083/1083.pdf.

[293] UK Parliament, 'Swearing in and the parliamentary oath', (website) https://www.parliament.uk/about/how/elections-and-voting/swearingin/, (accessed 22 March 2024).

[294] Committee on Standards, 'Oral evidence: House of Commons Standards Landscape, HC 247', Tuesday 6 February 2024, https://committees.parliament.uk/oralevidence/14230/pdf/.

Meanwhile on enforcement of standards, the momentum is building for change. There is clear public support. Ipsos UK and Full Fact's research has found that 81% of the UK public agrees with the statement that "it is important to hold public figures to a higher standard and demand truth from politicians".[295] Other organisations such as Unlock Democracy[296] and Transparency International[297] have called for independent oversight of the Ministerial Code and for it to be placed on a statutory footing.

These calls have made it into Parliament. Lord Anderson's recent private members' Public Service (Integrity and Ethics) Bill included putting the Ministerial Code on a statutory footing and giving an Independent Adviser statutory powers to investigate potential Code breaches and report on whether they'd occurred.[298]

Lord Anderson's Bill did not reach the debating stage, and the arguments against putting the Ministerial Code on a statutory footing should be acknowledged. The parliamentary secretary at the Cabinet Office, Alex Burghart MP, at the recent hearing of the Committee on Standards, explained that the status quo was important for accountability: "As a Minister, I am accountable to the Prime Minister; the Prime Minister is accountable to Parliament; and Parliament is accountable to the country."[299]

The implication of his remarks is that having independent oversight of the Ministerial Code would mean a move away from ensuring that the hiring and firing of ministers remain within the gift of the Prime Minister, and that in turn would mean a certain loss of the PM's authority over their government. The problem with this argument is that, following a hypothetical ministerial misdemeanour, and with a government majority in the Commons, it could take up to five years for the next time Parliament would be "accountable to the country". This is clearly far too slow for any kind of genuine accountability.

We must move forward. The public expectation is there, and the groundwork has been laid for what change might look like. Full Fact will work with campaign allies and parliamentarians following the general election on further proposals for reform. It's up to the new government and parliament to adopt these proposals, and commit to genuine reform of our political system. In an era when our information environment generates such uncertainty, and as AI and other new technologies accelerate that trend, greater transparency and accountability is the only way that public trust in politics can be restored.

[295] Ipsos, 'Full Fact UK Public Attitudes Research', April 2024, http://fullfact.org/audience-research-2023.

[296] Public Administration and Constitutional Affairs Committee, 'Written evidence from Unlock Democracy', November 2020, https://committees.parliament.uk/writtenevidence/15383/pdf/.

[297] Transparency International, 'It's time for the Ministerial Code to become law', 17 March 2021, https://www.transparency.org.uk/ministerial-code-UK-nolan-principles-public-ethical-standards.

[298] UK Parliament, 'Public Service (Integrity and Ethics) Bill', 31 October 2023, https://bills.parliament.uk/bills/3332.

[299] Committee on Standards, 'Oral evidence: House of Commons Standards Landscape, HC 247', Tuesday 6 February 2024, https://committees.parliament.uk/oralevidence/14230/pdf/.

Action for the government

- Place the Ministerial Code on a statutory footing, and ensure it has independent oversight.
- Legislate in a targeted way to end deceptive campaign practices, such as party. leaflets masquerading as newspapers.

Action for Parliament

- All new MPs must have training on standards within six months of their election, including guidance on sharing and presenting information accurately, and on how to pursue a correction from a fellow MP.
- MPs should be required to make a declaration to be honest and tell the truth. alongside swearing the oath to the King when they enter Parliament.

Action for political parties

- All parties should commit to independent regulation of their political advertising.

Full Fact
17 Oval Way
London
SE11 5RR

✉ fullfact.org/contact
🐦 @FullFact
🌐 fullfact.org

A registered charity (no. 1158683) and a non-profit company (no. 6975984) Limited by guarantee and registered in England and Wales