DEMOS

SYNTHETIC POLITICS

PREPARING DEMOCRACY FOR GENERATIVE AI

ELLEN JUDSON
SARAH A. FISHER
JEFFREY W. HOWARD
BEATRIZ KIRA
KIRAN ARABAGHATTA BASAVARAJ
HANNAH PERRY

In partnership with

UCL

MARCH 2024

Published by Demos March 2024

15 Whitehall, London, SW1A 2DD
T: 020 3878 3955
hello@demos.co.uk
www.demos.co.uk

CONTENTS

ACKNOWLEDGEMENTS

This project is a collaborative effort between Demos's CASM team and the UCL Digital Speech Lab.

Demos would like to thank UCL Digital Speech Lab for their generous support and partnership throughout this project. The UCL Digital Speech Lab is funded by UK Research & Innovation (grant reference MR/V025600/1); we thank UKRI for its support, which made this collaboration possible.

Thank you to the experts from across politics, academia, civil society and the technology sector for their valuable insights. The initial ideas for this policy briefing were presented at a workshop in London in January 2024, including:

- Representatives from DSIT
- Jessica Rose Smith and Laura Waters from Ofcom
- Nicola Aitken, Meta
- Niklas Eder, Meta Oversight Board
- Representatives from other anonymous social media platforms.
- Elliot Jones, Ada Lovelace
- Michael Tunks, IWF
- Dylan Sparks, Reset
- Alexandra Pardal, People vs Big Tech Coalition
- Felix Simon, Oxford Internet Institute
- Boxi Wu, Oxford Internet Institute
- Hannah O'Rourke, Campaign Lab

For relevant discussions on this and related work, we are grateful to Professor Marc Stears from the UCL Policy Lab, and the Policy team at the Electoral Commission.

For input into the final paper, we are grateful to Kyle Taylor and Felix Simon.

At Demos, we would like to thank our colleague, Hannah Perry, for the management of this project. Thanks also to Polly Curtis, Sophia Knight, Elizabeth Seger, Sumaya Akthar, Chloe Burke and Felix Arbenz-Caines.

Any errors remain the authors' responsibility.

Ellen Judson, Sarah A. Fisher, Jeffrey W. Howard, Beatriz Kira, Kiran Arabaghatta Basavaraj and Hannah Perry

March 2024

EXECUTIVE SUMMARY

Generative AI has taken the world by storm. Recent innovations in generative AI — AI systems that can produce synthetic text, image, video, and audio content — have enormous promise, enabling new forms of research and creative expression. But they also risk supercharging pre-existing risks, potentially unleashing harmful content on an unprecedented scale and with great impact. In a year in which more than half the world's population will head to the polls, and in which violent political conflicts intensify online debates, the stakes are incredibly high.

Public facing generative AI tools have the potential to change what and how content is created, and how it enters and spreads around the online world. These changes to the information environment have particular implications for democratic integrity: in the effects they have on core democratic ideals of equality, truth and non-violence in political discourse. How far-reaching these effects will be - and how much policy attention they should capture - is contested. In this paper we set out proportional recommendations to mitigate risks and maximise opportunities of generative AI, while also supporting a broader healthier information environment.

In **Action Plan 1**, we consider the actions that should be urgently put in place to reduce the acute risks to democratic integrity presented by generative AI tools in the context of this year's remaining global elections. These risks include enabling more effective gendered and racialised disinformation campaigns, exacerbating distrust in elections, and enabling the fomentation of civil unrest. We set out recommendations to (1) reduce the production and dissemination of harmful synthetic content and (2) to empower users so that harmful impacts of synthetic content are reduced in the immediate term.

In **Action Plan 2**, we set out a longer-term vision for how the fundamental risks to democratic integrity should be addressed. We set out the ways in which generative AI tools can help bolster equality, truth and non-violence, from enabling greater democratic participation to improving how key information institutions operate. Before this positive potential for AI can come to fruition, however, it is essential that fundamental threats of bias, inaccuracy and opacity in generative AI systems are overcome. Finally, we consider ways in which tech companies and policymakers can work together to improve the quality of an AI-driven information environment, empower citizens in democracy, and develop generative AI tools to serve the public interest.

RECOMMENDATIONS

Our recommendations for public policymakers and regulators are primarily focused on the UK. Even so, the general principles of truth, equality and non-violence we defend are relevant to other jurisdictions at different stages of digital policymaking, such as the US and EU.

TABLE 1
ACTION PLAN 1: IMMEDIATE ACTIONS TO TACKLE DEMOCRATIC RISKS

WHO	RECOMMENDATIONS	REFERENCE IN MAIN REPORT[1]
AI Companies	Developers of generative AI foundation models and user applications should develop and **publish much more specific policies concerning the content that users may and may not generate**, especially with respect to content that undermines democratic integrity. These policies should be the explicit basis of models' guardrails, creating a symmetry between expectations on users and on the company itself.[2]	**AP1.1**
	Companies should **upscale "prompt-hacking" and "red-teaming" exercises ahead of elections** to help identify and mitigate model misuse, and should publish summaries of what they are doing in this regard.	**AP1.2**
	Developers of generative AI foundation models and user applications should **watermark the contents produced by their tools**, where it is feasible in the short term.[3] This is likely to predominantly be feasible for the largest companies, many of whom already have watermarks embedded in their tools.	**AP1.7**
	Developers of generative AI applications for text generation should **ensure their tools provide clear information to users about the potential inaccuracy** of the content produced, with an explanation that these tools are not reliable sources of factual information.	**AP1.8**
Social media platforms	**Rather than create new *sui generis* rules for synthetic content, platforms should double down on the enforcement of rules** against harmful speech and rules about advertising on their platforms for all users, removing content that breaches their policies regardless of whether it is generated by human or machine.	**AP1.3**
	Content-distribution platforms should **require labelling synthetic user-generated content and ads**, including by automatically labelling that which is produced by their own in-house tools, and by enabling users to label their posts or suggest the labelling of other posts.[4]	**AP1.6**
	Ahead of elections, platforms should: a. **Ensure there are transparent escalation systems** and clear channels of communication in place for those targeted to report harassment campaigns.[5] b. **Ensure signposts are easily available to resources for further support** for those targeted.[6]	**AP1.9**

1 References refer to where this recommendation falls in the detailed Action Plans in the main body of the paper. For example, AP1.1 is Action Plan 1 Recommendation 1.
2 E.g. OpenAI has begun this work. Open AI, January 2024. https://openai.com/blog/how-openai-is-approaching-2024-worldwide-elections
3 See the Content Authenticity Initiative's work on this. https://contentauthenticity.org/how-it-works
4 E.g. Meta's approach. CNBC, November 2023. https://www.cnbc.com/2023/11/28/meta-updates-political-ad-rules-to-cover-ai-generated-images-videos.html
5 Judson,E. July 2021 https://eu.boell.org/en/2021/07/09/gendered-disinformation-6-reasons-why-liberal-democracies-need-respond-threat. UNESCO, November 2023. https://www.unesco.org/en/articles/technology-facilitated-gender-based-violence-times-generative-ai
6 Demos, Threats Women Journalists Face Online: Ana's Story and Samina's Story: https://www.youtube.com/channel/UCeO4Yd8qi4y4IsjrdYHhJQQ/videos; Digital Rights Foundation, 2022 https://digitalrightsfoundation.pk/wp-content/uploads/2023/07/Policy-Brief-Cyber-Harassment-Helpline-2022.pdf; Deelen & Voght, February 2024. https://www.irex.org/insight/what-media-and-civil-society-leaders-can-do-mitigate-technology-fueled-misogyny-2024

Social media platforms	Platforms should **make data available to researchers** to support independent research into the effects of AI-generated content and countermeasures such as labelling on the spread of deceptive content on their services.	**AP1.11**
UK policymakers	Policymakers should invest in independent research to **monitor the volume, type and potential effects of synthetic content generated in the run-up to the current set of elections as far as possible**, as well as the risks it poses to the democratic principles of Truth, Equality, and Non-violence, with findings to inform the development of effective regulatory oversight further down the line *(see Action Plan recommendation 2.13)*.	**AP1.4**
Regulators and law enforcement	Regulators and law enforcement should **assess the extent to which they could mitigate democratic risks acting within their existing mandates and issue specific guidance** clarifying how existing law and policy/regulation already applies to generative AI and the democratic risks outlined here. (See, for instance, the FCC in the US taking rapid action against AI-generated robocalls).[7]	**AP1.5**
Political parties	Political parties should **develop a cross-party compact on how generative AI is to be used transparently and ethically in election campaigning.** This should include a commitment to not amplifying content about any candidate or party that there are reasonable grounds to believe is materially deceptive.[8]	**AP1.10**

TABLE 2
ACTION PLAN 2: LONG-TERM STRATEGIES TO PROTECT DEMOCRATIC INTEGRITY

WHO	RECOMMENDATIONS	REFERENCE IN MAIN REPORT
PROTECTING DEMOCRATIC INTEGRITY FROM AI		
Social media platforms	Content-distribution platforms should **conduct and publish risk assessments of the integration of generative AI tools into their services** before they are integrated. (Ideally, this would form part of their duties under the Online Safety Act).	**AP2.17**
AI Companies	Companies producing text generation tools should **explore how to build reliable citations into their search results**, enabling generative AI tools to provide links to reliable sources that can be independently checked for any apparently factual information they produce, rather than only relying on content warnings about inaccuracy.[9]	**AP2.14**

7 BBC, February 2024. https://www.bbc.co.uk/news/world-us-canada-68240887 ; Content Authenticity, February 2024 https://contentauthenticity.org/blog/february-2024-this-month-in-generative-ai-election-season
8 Demos, January 2024. https://demos.co.uk/wp-content/uploads/2024/01/Generating-Democracy-Report-1.pdf
9 La Trobe University. https://latrobe.libguides.com/artificial-intelligence/referencing

AI Companies	Developers of AI models and applications should **put significant resources into understanding — and explaining — the provenance of AI-generated inaccuracies and biases, and take meaningful steps to rectify these** (e.g. through better curated training data, more human feedback, or more sophisticated guardrails).[10] This should broaden out from the scope of the Munich Accord[11] to cover harms to truth, equality and non-violence, and not only deceptive election content.	**AP2.16**
	AI tool producers should **design the interfaces of their products and services to effectively communicate their product purpose and limitations to users.**	**AP2.20**
Both AI companies and social media platforms	AI companies and social media platforms should **work together to deploy more interoperable watermarking solutions**, such as are being developed through the Content Authenticity Initiative technical standards, more robust, 'maximally indelible' watermarks and disclosure of the AI-generation of content.[12]	**AP2.12**
Policymakers, AI companies and social media companies	Stakeholders should continue to **support independent research, through means such as funding and data access provision, to identify the lessons learnt from ongoing monitoring during the election period**, and integrate these learnings into the deployment of future safeguards against democratic risk.	**AP2.13**
UK policymakers	UK policymakers should **impose obligations on AI companies requiring them to undertake comprehensive risk assessments, with a focus on the risks that their models and products pose to democratic integrity.** This should be enforced through meaningful audit by regulators and routes to data access for independent civil society organisations.[13] The newly-passed EU AI Act moves in this direction, with duties on developers of general-purpose AI models with systemic risks (which can include models used for generative AI tools) to assess and mitigate these risks,[14] while the Digital Services Act provides for data access to social media platform data: but the UK is far behind.	**AP2.19**
Policymakers, regulators, and civil society oversight bodies	Stakeholders should **assess AI and social media companies for their efficacy post-election with regard to the actions recommended in Action Plan 1.** Such recommendations should also inform the regulatory duties and codes of practice that companies will be required to abide by and report against (e.g. codes of practice drafted by Ofcom in enforcing the Online Safety Act).	**AP2.15**
All stakeholders	Industry standards should be **set within sectors through collaboration between companies, regulators and civil society organisations to define sector-leading usage rules and best practice for generative AI tools**, which companies could then be certified on the basis of their compliance with.[15]	**AP2.18**

10 E.g. Mitchell, February 2024. https://time.com/6836153/ethical-ai-google-gemini-debacle/
11 AI Elections Accord. 2024. https://securityconference.org/en/aielectionsaccord/
12 E.g. the duties in the EU AI Act around transparency of outputs and disclosures of AI uses, as well as, https://contentauthenticity.org/; https://spectrum.ieee.org/meta-ai-watermarks; E.g. BBC, March 2024. https://www.bbc.co.uk/rd/blog/2024-03-c2pa-verification-news-journalism-credentials#:~:text=Like%20an%20audit%20trail%20or,where%20it%20has%20come%20from.
13 Algorithm Watch, December 2022. https://algorithmwatch.org/en/dsa-data-access-explained/
14 European Parliament, March 2024. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html
15 City AM, February 2024. https://www.cityam.com/the-notebook-the-city-has-to-unite-against-the-risks-of-generative-ai/; Demos, January 2024. https://demos.co.uk/research/generating-democracy-ai-and-the-coming-revolution-in-political-communications/

PROMOTING DEMOCRATIC INTEGRITY WITH AI		
AI companies	Companies should **publish the principles on which their AI tools have been designed and trained**, how this has been achieved and with what oversight (e.g. through 'democracy-by-design' procedures, training procedures, and independent oversight or audit procedures).[16]	**AP2.22**
Funders	Funding bodies in the public and private sector should **further incentivise the development of democratically beneficial generative AI applications.**[17] This might be achieved, for example, by supporting a 'Democratic Sandbox' for companies, civic tech and civil society organisations to collaborate and experiment with developing public and open democratic AI systems or by supporting the development of generative AI tools and best practice for public interest functions such as charities and public interest news organisations.[18]	**AP2.21**
Policymakers	Policymakers should **engage with the public deliberations on governance of generative AI** and support these to be scaled and implemented into policymaking processes.[19]	**AP2.24**
UK regulators	Regulators should **collaborate to produce consistent guidance for the development of industry best practice in use of generative AI.** Regulators should set out this intention in their upcoming strategic guidance requested by the UK government to be published at the end of April 2024.[20]	**AP2.23**

16 E.g. DSIT, February, 2024. https://assets.publishing.service.gov.uk/media/65ccf508c96cf3000c6a37a1/Introduction_to_AI_Assurance.pdf
17 https://www.aielectionsaccord.com/uploads/2024/02/A-Tech-Accord-to-Combat-Deceptive-Use-of-AI-in-2024-Elections.FINAL_.pdf
18 Linklaters, February 2024. https://techinsights.linklaters.com/post/102izns/prepare-for-take-off-uks-digital-regulatory-cooperation-forums-ai-and-digital; https://huit.harvard.edu/ai-sandbox; DSIT, August 2023. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper; Taylor, K, February 2024. https://www.youtube.com/watch?v=MHgL9PZxits ; GMF, February 2024. https://www.gmfus.org/news/gmf-launches-project-pioneer-novel-technologies-strengthen-democratic-resilience; House of Lords, 2023.https://bills.parliament.uk/publications/53068/documents/4030; Ajder, H, July 2023. https://www.linkedin.com/posts/henryajder_a-new-national-purpose-ai-promises-a-world-leading-activity-7075041764121661440-bE6J/
19 Hono, S.Y., February 2024. https://openfuture.eu/blog/alignment-assembly-on-ai-and-the-commons/; Belgium24, February 2024. https://belgian-presidency.consilium.europa.eu/en/news/launch-of-citizens-panel-on-artificial-intelligence/
20 DSIT, 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#fn:28

INTRODUCTION

This year is a politically momentous one, with almost half the world voting in elections.[21] The injection of generative AI into the public domain brings with it the potential to transform our information environments and political discourse by making them more effective, more relevant, and more participatory. At the same time, risks abound: that they will become more manipulative, more confusing, and more dangerous. Synthetic content produced by generative AI poses risks to core democratic values of truth, equality, and non-violence—substantially exacerbating problems that have afflicted our digital ecosystem over the past 10 years. The question, then, is what private and public decision-makers can do to reduce those risks.

In **Action Plan 1**, we set out actions that AI companies, social media platforms, and UK policymakers should be doing already to reduce the democratic harms likely to arise acutely in the context of a highly politicised electoral year. These include demands for AI companies to enact more powerful safety standards for content production and model use , and for social media platforms to improve labelling of synthetic content while doubling down on enforcing their existing misinformation policies.

However, there are limitations on what can be achieved this year to safeguard imminent elections; so it is important that longer term plans are also conceived now.

In **Action Plan 2**, we set out longer-term systemic solutions, offering a vision toward which decision-makers should aspire after the many 2024 elections have passed. These solutions will take longer to implement, but with the payoff that they help tackle democratic challenges at a more fundamental level. As part of that vision, we also consider the ways in which generative AI might be improved so that it can constitute a force that *bolsters* democratic integrity instead of undermining it.

We focus on challenges to democratic discourse. Our recommendations to global technology companies for preserving truth, equality, and non-violence are, in principle, applicable internationally. We acknowledge, though, that there are distinctive challenges and risks to citizens in authoritarian regimes that merit a separate, focused treatment, beyond the scope of what we offer here.

Our recommendations for public policymakers and regulators are primarily focused on the UK. Even so, the general principles we defend are relevant to other jurisdictions at different stages of digital policymaking, such as the US and EU.

21 Ewe, K., December 2023. https://time.com/6550920/world-elections-2024/

PRINCIPLES OF AND RISKS TO DEMOCRATIC INTEGRITY

PRINCIPLES OF DEMOCRATIC INTEGRITY

In this 'year of elections', generative AI policy discussions increasingly focus on the risks to elections themselves, with the defence of 'electoral integrity' a key concern. For instance, many have worried that people's votes could be swayed on the day by an AI-driven, highly effective, foreign disinformation campaign.[22] But this focus is too narrow.[23] In this paper, we focus instead on the much broader value of *democratic integrity.*[24] The changes brought by generative AI have wide potential impacts across the core values that a democratic society depends on - of equality, truth, and non-violence. These are a minimum prerequisite for democratic processes and institutions - from elections to the operations of government - to be able to operate with integrity. It is these ideals that may be endangered by generative AI unless it is properly governed. New technologies should defend and promote these values, and not put them under further strain.

Why truth, equality and non-violence?

A democracy must live up to its underlying core ideals: the protection of equality, truth and non-violence. We derive these ideals from a familiar understanding of a democracy as a political association of free and equal citizens who govern themselves through reason.[25] Although citizens may disagree about which policies are best, they work through those disagreements through public conversation. This public discourse must be genuinely equally accessible, so that people can participate freely and authentically, and different interests weighed equally. The complexity of public policy, and the scale of our societies, mean that citizens' deliberations also include discussion of legislative candidates, chosen through voting in elections. Citizens in such a democracy respect the rights of all to vote and stand for office, and respect the outcomes of electoral

22 Ambrose, T, 2024. https://www.theguardian.com/uk-news/2024/feb/25/uks-enemies-could-use-ai-deepfakes-to-try-to-rig-election-says-james-cleverly

23 Further research is also needed to evaluate the size of the risk posed by AI to the perceived legitimacy of a particular electoral vote.

24 See the Electoral Commission: 'We work to promote public confidence in the democratic process and ensure its integrity'. https://www.electoralcommission.org.uk/about-us#:~:text=The%20Electoral%20Commission%20is%20the,process%20and%20ensure%20its%20integrity

25 For some classic texts on deliberative democracy, see Gutmann, A. and Thompson, 2004, Why Deliberative Democracy? (Princeton: Princeton University Press); Rawls, J., 2005, "The Idea of Public Reason Revisited," in Political Liberalism, expanded edition (New York: Columbia University Press), Habermas, J. 1992, Between Facts and Norms: A Discourse Theory of Law and Democracy (Cambridge, MA: MIT Press).

decisions, without threats of violence. This, anyway, is a familiar and powerful vision of a democracy, and it is our starting point here.

The ideals that help to constitute this vision of democratic integrity are multiple, and there are of course many institutional and procedural requirements that democracies must meet. Here we focus on the three fundamental principles, without which further democratic principles and procedures cannot legitimately function.[26]

- **Equality.** We are committed to a society in which citizens regard one another as equals, affirming one another's equal right to participate and to be treated with respect.
- **Truth.** We are committed to a public conversation in which citizens and their representatives sincerely deliberate, with a shared understanding of basic facts and criteria for evaluating the truth of claims.
- **Non-violence.** We are committed to managing our disagreements through respectful engagement and voting instead of the use of force, respecting the outcomes of elections and the peaceful transition of power.

Our focus is on how these principles may be upheld or undermined through public discourse and information environments, as this is where citizens take part in political discussions; and where bad actors may seek to disrupt democratic discussions through information warfare.[27] As such, we discuss the risks and opportunities arising from public-facing generative AI tools which are most likely to impact these principles by affecting what content and information is produced, shared and consumed. These tools include text generators like ChatGPT and image generators such as Midjourney, as well as upcoming video generators such as Sora.

We consider what obligations fall on companies who develop and release these tools, which are primarily used by citizens to access or produce information (even as they have other purposes). We also consider social media platforms where such content may be shared, disseminated or amplified. These (increasingly overlapping) actors—those designing and releasing generative AI products, and those hosting content produced by those tools—are the primary addressees of our argument. We will also pinpoint what policymakers and regulators should do to hold companies to account in living up to their obligations.

RISKS TO DEMOCRATIC INTEGRITY

Having articulated our focus on democratic integrity (beyond mere electoral integrity) and three core democratic ideals, we now ask: how might public generative AI tools pose a risk to their realisation? We are already familiar with the multifarious ways in which online content can undermine democratic integrity; after all, this has been the central complaint about social media for over a decade. *Equality* has been undermined through hateful online campaigns involving harassment, bullying, and other abusive speech. *Truth* has been under attack from varieties of mis- and dis-information. *Nonviolence* has been subverted through speech inciting and threatening violence against political opponents, and in attempts to thwart the successful transfer of power.

The immediate problem with generative AI, then, is not that it unleashes a completely new set of harms with which we are wholly unfamiliar (though it may produce new harms in the future). The pressing risk, instead, is that it leverages the computational power of advanced machine learning systems to supercharge pre-existing problems—making it cheaper to produce and propagate harmful content, making much of this content more impactful than it would otherwise be, and (increasingly) powering creative forms of manipulation.[28] To be sure, a great deal of synthetic content is innocuous, or even beneficial, at least in its effects on audiences.[29] Our concern is with the subset that is decidedly neither.

26 These values have corollaries enshrined in regulation in the UK: for instance, the Equality and Human Rights Commission (EHRC) affirms their role in 'enforcing and upholding the laws that safeguard everyone's right to fairness, dignity and respect', and the Equality Act 2010 makes it unlawful to discriminate against or harass individuals based on protected characteristics.

27 Jungherr, A. 2023. https://journals.sagepub.com/doi/10.1177/20563051231186353

28 A Lords Committee report on LLMs notes that 'the most immediate security concerns from LLMs come from making existing malicious activities easier, rather than qualitatively new risks,", though concludes that the threat of disinformation, hallucinations, deepfakes means that a 'reasonable worst case scenario' would be the integrity of the election being called into question. Communications and Digital Committee, February 2024. https://publications.parliament.uk/pa/ld5804/ldselect/ldcomm/54/5402.htm , p.39 and p.42

29 There are, to be sure, distinct concerns about data privacy or intellectual property; these surely matter, but they are not democratic concerns.

We are already seeing signs that generative AI is affecting our democratic politics. In Slovakia, pundits speculate whether a deepfake about a politician two days before the national vote ended up swaying the election.[30] In Pakistan, an imprisoned Imran Khan was unable to communicate with his voters, so his team deployed a deepfake instead.[31] In Argentina's elections, AI-generated images have been ubiquitous.[32] In the U.K., manipulated audio of Keir Starmer depicted him berating and cursing at his team.[33] In the Republican primary in New Hampshire, nearly 25,000 "robocalls" to voters used faked audio of President Biden enjoining them not to vote.[34] And in the Indonesian election, Midjourney and ChatGPT were deployed in Prabowo Subianto's campaign to create campaign imagery and send tailored messages to voters.[35]

The potential for generative AI to undermine democracy is significant: from the ability of generative AI to produce realistic audio and video which can effectively deceive or confuse voters at scale, to risks of violence.[36] The World Economic Forum's report on global risks ranked dis- and mis-information as the greatest short-term risks.[37] And even just the awareness of these risks, experts have warned, means that citizens may be less trusting of reliable information and be more easily swayed by claims of AI-based electoral interference.[38]

However, many of the concerns are still theoretical and beyond anecdotes, not yet widely borne out.[39] The likelihood of these risks eventuating in actual election interference, information chaos or widespread real-world harm, is greatly contested.[40] Other risks to democracy persist, meaning that although we may see greater AI-powered disinformation, this is unlikely to greatly worsen our existing information disorder, which is already driven by powerful actors successfully using their platforms and conventional tools.[41,42] Experts warn that AI panic, in an already fraught year for political trust and stability, could do more harm than good: especially if it means that attention is focused on high-profile generative AI cases rather than the existing information threats causing the widest risk.[43]

If we overplay the risks of AI, we risk playing into the easily weaponized narrative that our democratic institutions are fragile,[44] and focus limited political will on bringing in ineffective and potentially harmful policies to tackle a disinformation bogeyman.[45] If we underplay them, however, we risk losing the window in which to take preventive action to safeguard our democratic processes and principles against the exploitation of AI by bad actors[46] - a common failing in digital policymaking.[47] There is much we still do not know about

30 Solon, O., September 2023. https://www.bloomberg.com/news/articles/2023-09-29/trolls-in-slovakian-election-tap-ai-deepfakes-to-spread-disinfo?leadSource=uverify%20wall;
31 Folkman, V., February 2024. https://www.politico.eu/article/pakistans-imran-khan-use-ai-artificial-intelligence-make-victory-speech-from-jail/
32 NYTimes, November 2023, https://www.nytimes.com/2023/11/15/world/americas/argentina-election-ai-milei-massa.html
33 FullFact, October 2023. https://fullfact.org/news/keir-starmer-audio-swearing/
34 Tolan, C, O'Sullivan, D. and Winter, J. February 2024. https://edition.cnn.com/2024/02/07/politics/biden-robocall-texas-strip-mall-invs/index.html; Borron-Lopez, L. and Popat, S, February 2024. https://www.pbs.org/newshour/show/how-ai-generated-misinformation-threatens-election-integrity BBC, February 2024. https://www.bbc.co.uk/news/world-us-canada-68240887
35 Duffy, K., February 2024. https://www.cfr.org/blog/ai-context-indonesian-elections-challenge-genai-policies
36 Such risks are discussed in more detail below. See also: Yusypovych, S., February 2024. https://www.linkedin.com/pulse/reflection-assessing-ai-borne-risks-integrity-2024-us-yusypovych-e1cpe/; Van Der Linden, S., January 2024. https://www.wired.com/story/ai-generated-fake-news-is-coming-to-an-election-near-you/; Gorman, L. and Levine, D. February 2024. https://securingdemocracy.gmfus.org/wp-content/uploads/2024/02/The-ASD-AI-elections-Security-Handbook_.pdf
37 World Economic Forum, January 2024. https://www.weforum.org/press/2024/01/global-risks-report-2024-press-release/
38 Gorman, L. and Levine, D. February 2024. https://securingdemocracy.gmfus.org/wp-content/uploads/2024/02/The-ASD-AI-elections-Security-Handbook_.pdf
39 Simon, F., Altay, S., Mercier, H., October 2023. https://misinforeview.hks.harvard.edu/article/misinformation-reloaded-fears-about-the-impact-of-generative-ai-on-misinformation-are-overblown/; Wirtschafter, V. January 2024. https://www.brookings.edu/articles/the-impact-of-generative-ai-in-a-global-election-year/
40 Ajder, H. February 2024. https://www.linkedin.com/feed/update/urn:li:activity:7167915442475368448/
41 Law, H., January 2024. https://www.learningfromexamples.com/p/the-marginal-risk-of-ai
42 Nielsen, R. January 2024. https://www.ft.com/content/5da52770-b474-4547-8d1b-9c46a3c3bac9; Marcus, P. February 2024. https://www.linkedin.com/posts/petemarcus82_todays-times-headline-is-a-great-example-activity-7167780930454020096-H0fe?utm_source=share&utm_medium=member_desktop
43 Rashid Diya, S. February 2024. https://www.context.news/ai/opinion/cheap-fakes-are-a-blind-spot-for-platforms-in-the-global-south; Marcus, P. February 2024. https://www.linkedin.com/feed/update/urn:li:activity:7167780930454020096/
44 Gorman, L and Levine, D., February 2024. https://securingdemocracy.gmfus.org/wp-content/uploads/2024/02/The-ASD-AI-elections-Security-Handbook_.pdf
45 Center for News, Technology & Innovation, February 2024. https://innovating.news/article/most-fake-news-legislation-risks-doing-more-harm-than-good-amid-a-record-number-of-elections-in-2024/
46 Council on Foreign Relations, December 2023. https://www.cfr.org/podcasts/year-ai-and-elections
47 This uncertainty about how best to protect democracy mirrors the wider debate about AI - whether we should focus on preventing possible long-term existential risks, or immediate, real-world harms caused by AI: and whether focusing on one means the other will be inevitably overlooked by policymakers. Hanna, A. and Bender, E.M. August 2023. https://www.scientificamerican.com/article/we-need-to-focus-on-ais-real-harms-not-imaginary-existential-risks/; Farrell, H., December 2023. https://www.economist.com/by-invitation/2023/12/12/ais-big-rift-is-like-a-religious-schism-says-henry-farrell

the capabilities or uses of these technologies, and we must be willing to adapt our responses as new evidence comes to light along the way.[48]

In the midst of uncertainty, in this paper we seek to offer a middle way, navigating between the horror of AI doom and the comfort of complacency. Our recommendations are intended to be proportional to the level of risk: suggesting low-lift proposals that could reduce risks without significant negative consequences, and more substantial interventions for much higher risks. We include recommendations for challenging the harms of synthetic content which would also have positive ramifications for reducing risks of other information harms.

48 With thanks to Felix Simon, pers. comm. 2024

ACTION PLAN 1
URGENT ACTIONS TO REDUCE ACUTE RISKS TO DEMOCRATIC INTEGRITY IN 2024

Generative AI threatens to increase extant risks to democratic integrity at a critical moment in the electoral cycle. Despite the inherent uncertainty in many of these risks and what countermeasures will necessarily be the most effective, it is reasonable to demand that relevant actors work to reduce them. We will discuss risks to each of the democratic values raised above: equality, truth, and nonviolence. We'll set out the current action being undertaken by key players, and where immediate improvements could be made within the next few months. These do not constitute a comprehensive plan to *eliminate* risk, but a minimum standard of action that actors can reasonably be asked to pursue.

RISKS TO EQUALITY

Generative AI risks supercharging the production of various forms of hateful speech, understood as content that attacks the equal standing of all citizens (usually, the equal standing of historically oppressed groups). Hateful speech often takes the form of disinformation, spreading defamatory lies against groups and their members.[49]

Generative AI poses a risk that such identity-based disinformation attacks targeting marginalised groups are made easier and more scalable. Such attacks potentially may also be more credible[50] and invasive.[51] The use of generative AI tools to create deepfakes (i.e. fabricated audiovisual content) is a particularly noxious mechanism of gendered and racialised disinformation campaigns. For instance, women politicians and journalists are disproportionately targeted through the creation and sharing deepfakes of intimate images.[52]

49 This is central to the analysis of hate speech in Waldron, The Harm in Hate Speech (Cambridge, MA: Harvard University Press, 2012).
50 Huschens, M. et al, September 2023. https://arxiv.org/abs/2309.02524
51 Chowdhury, R. and Lakshmi, D. 2023. https://unesdoc.unesco.org/ark:/48223/pf0000387483
52 Koltai, K., November 2023. https://www.bellingcat.com/news/2023/11/27/anydream-secretive-ai-platform-broke-stripe-rules-to-rake-in-money-from-nonconsensual-pornographic-deepfakes/

These can then be spread with the aim of demonising, sexualising or humiliating their targets[53,54,55] - and disproportionately affecting Black women and women from other minoritised groups.[56] Among its various harms, such content affronts democratic equality by attacking the equal status of these citizens and making it more difficult for them to participate in public life.

Specific risks to *Equality* include:

RISK	EXAMPLE
Synthetic content being used to harass political candidates/their supporters.	Non-consensual intimate deepfakes of women in public life being shared - creating pornographic or private images.[57,58] Chatbots offering highly personalised how-to templates for harassment.[59] Gendered and racialised disinformation campaigns becoming supercharged.[60]
Synthetic content invoking harmful stereotypes.	Racial stereotypes being reproduced in AI-generated imagery.[61]

RISKS TO TRUTH

Generative AI risks supercharging the production of various forms of misinformation, including malicious disinformation, about political issues, candidates, and processes. One risk is that generative AI will produce, or be maliciously weaponized to produce, convincing falsehoods.[62] Content that is partly true but nevertheless highly misleading is usually even more convincing.[63] Such misinformation undermines *Truth* by inhibiting citizens from sharing a common reservoir of factual information to underpin their public deliberation– especially (but not only) in the run-up to elections.[64]

53 Jankowicz, N., June 2023. https://www.theatlantic.com/ideas/archive/2023/06/deepfake-porn-ai-misinformation/674475/
54 Posetti, J. and Shabbir, N., November 2022. https://www.icfj.org/sites/default/files/2022-11/ICFJ_UNESCO_The%20Chilling_2022_1.pdf
55 McGlynn, C. and Rackley, E., May 2021. https://www.claremcglynn.com/_files/ugd/e87dab_b3a67112fc76434dba953514053c8152.pdf
56 Glitch UK, 2023. https://glitchcharity.co.uk/wp-content/uploads/2023/07/Glitch-Misogynoir-Report_Final_18Jul_v5_Single-Pages.pdf; Overton Testimony, November 2023. https://oversight.house.gov/wp-content/uploads/2023/11/Overton-Testimony-on-Advances-in-Deepfake-Technology-11-8-23-1.pdf McGlynn, C. July 2023. https://theconversation.com/if-someone-posts-your-private-photos-online-there-has-been-little-you-can-do-about-it-how-changes-in-the-law-will-finally-help-victims-209048; Glitch, 2023. https://glitchcharity.co.uk/wp-content/uploads/2023/07/Glitch_ENAR-Workshop-1-briefing.pdf
57 Jankowicz, N. December 2017. https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women/ ; Occenola, P. December 2018. https://www.rappler.com/newsbreak/in-depth/217563-disinformation-gone-macho/ Jankowicz, N. June 2023. https://www.theatlantic.com/ideas/archive/2023/06/deepfake-porn-ai-misinformation/674475/ Posetti, J. and Shabbir, N., November 2022. https://www.icfj.org/sites/default/files/2022-11/ICFJ_UNESCO_The%20Chilling_2022_1.pdf McGlynn, C. and Rackley, E., May 2021. https://www.claremcglynn.com/_files/ugd/e87dab_b3a67112fc76434dba953514053c8152.pdf
58 Glitch. 2023. https://glitchcharity.co.uk/wp-content/uploads/2023/07/Glitch-Misogynoir-Report_Final_18Jul_v5_Single-Pages.pdf ; McGlynn, C. July 2023. https://theconversation.com/if-someone-posts-your-private-photos-online-there-has-been-little-you-can-do-about-it-how-changes-in-the-law-will-finally-help-victims-209048; Glitch, 2023. https://glitchcharity.co.uk/wp-content/uploads/2023/07/Glitch_ENAR-Workshop-1-briefing.pdf
59 UNESCO, 2023. https://unesdoc.unesco.org/ark:/48223/pf0000387483/PDF/387483eng.pdf.multi
60 Demos, October 2020. https://demos.co.uk/research/engendering-hate-the-contours-of-state-aligned-gendered-disinformation-online/
61 Luccioni et al. November 2023. https://arxiv.org/abs/2303.11408 Nicoletti, L. and Bass, D. Date Unknown. https://www.bloomberg.com/graphics/2023-generative-ai-bias/
62 Simon, F., Altay, S., Mercier, H., October 2023. https://misinforeview.hks.harvard.edu/article/misinformation-reloaded-fears-about-the-impact-of-generative-ai-on-misinformation-are-overblown/
63 Allen, J et al. October 2023. https://jenny-allen.com/publication/allen-2023-vaccine/
64 Morrison, M. and Raj Singh, S. February 2024. https://foreignpolicy.com/2024/02/09/social-media-tech-election-disinformation-war/

Specific risks to *Truth* include:

RISK	EXAMPLE
False or misleading synthetic user generated content or ads about election processes, including when/where/how to vote (especially concerning if scaled).	Chatbot giving wrong information in response to queries about elections.[65] Spread of synthetic content or ads purporting to show that the election was 'rigged' (particularly in the US).[66] Chatbots 'hallucinating' - creating highly plausible information which is presented as factual but is false and has no basis in reality.
False or misleading synthetic content or ads about political candidates or parties, either to undermine or bolster support for them (which could have significant effect if viral content occurs at key moments in electoral cycle).	Deepfakes purporting to show politicians engaging in abusive behaviour.[67] Deepfakes showing politicians confessing to crimes.[68] Deepfakes of politicians doing positive things to garner greater support, such as fake meetings with constituents of a certain demographic.[69]
False or misleading synthetic content or ads about topics on which elections are being fought.	Non-existent newspaper articles referenced by chatbots.[70] Deepfakes of journalists.[71]
Targeted campaigns that serve different facts to different individuals / groups.[72]	Using generative AI to generate different adverts to better persuade people with different personality traits.[73]
Widespread distrust or scepticism in response to pervasive synthetic content.[74]	Growing distrust of elections and officials.[75]

65 Angwin, J. et al. February 2024. https://www.proofnews.org/seeking-election-information-dont-trust-ai/; For example, in one case Chatbots gave wrong information 30% of the time in responded to queries about European elections: https://www.washingtonpost.com/technology/2023/12/15/microsoft-copilot-bing-ai-hallucinations-elections/

66 With thanks to Kyle Taylor, pers. comm., 2024; https://www.theguardian.com/technology/2023/nov/15/facebook-ads-2020-election-rigged-stolen-instagram-policy

67 For example, an audio deepfake depicted Keir Starmer cursing abusively at his staff: Sky News, October 2023. https://news.sky.com/story/labour-faces-political-attack-after-deepfake-audio-is-posted-of-sir-keir-starmer-12980181

68 For example, a deepfake of a Slovakian political party leader involved him "confessing" to electoral fraud just days before the election, which his rivals won: Meaker, M. October 2023. https://www.wired.co.uk/article/slovakia-election-deepfakes Farid, H. https://farid.berkeley.edu/deepfakes2024election/

69 Farid, H. January 2024. https://contentauthenticity.org/blog/february-2024-this-month-in-generative-ai-election-season; Sharma, S. March 2023. https://www.independent.co.uk/news/world/americas/us-politics/donald-trump-ai-praying-photo-b2307178.html; Spring, M. March 2024. https://www.bbc.co.uk/news/world-us-canada-68440150

70 For example, ChatGPT has hallucinated Guardian articles that never existed: Moran, C. April 2023. https://www.theguardian.com/commentisfree/2023/apr/06/ai-chatgpt-guardian-technology-risks-fake-article

71 Solymos, K.K. October 2023. https://ipi.media/slovakia-deepfake-audio-of-dennik-n-journalist-offers-worrying-example-of-ai-abuse/

72 Simchon, A. et al. February 2024. https://academic.oup.com/pnasnexus/article/3/2/pgae035/7591134?login=false; Fredheim, R. and Pamment, R. February 2024. https://link.springer.com/article/10.1057/s41254-023-00322-5

73 For a study of the potential use and effectiveness of generative AI in political micro-targeting, see Simchon, A. et al (2024). https://doi.org/10.1093/pnasnexus/pgae035.

74 Gorman, L and Levine, D., February 2024. https://securingdemocracy.gmfus.org/wp-content/uploads/2024/02/The-ASD-AI-elections-Security-Handbook_.pdf Past experience suggests that this is likely to be a particular concern with respect to women in public life. See, for example: Jankowicz, N. December 2017. https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women/; Occenola, P. December 2018. https://www.rappler.com/newsbreak/in-depth/217563-disinformation-gone-macho/.

75 Smith, C. October 2023. https://www.governing.com/security/election-integrity-in-the-age-of-generative-ai-fact-vs-fiction

RISKS TO NON-VIOLENCE

Generative AI also poses a threat to the value of *Non-violence*. It can be used to produce *targeted threats of violence*, seeking to intimidate people from voting, seeking public office, or expressing their point of view. And it can also be used to produce effective *incitements to violence*–especially pernicious in the context of contested elections, where the risks of street violence are high.

Specific risks to *Non-violence* include:

RISK	EXAMPLE
Synthetic content suppressing voting through intimidation.	Robocalls seeking to suppress or intimidate voters imitating a credible source.[76]
Synthetic content inciting violent political protest.	Mobs protesting legitimacy of an election.[77]
Synthetic content glorifying violence or dangerous organisations.	Chatbot producing content using themes found in extremist propaganda.[78] Chatbots designed specifically to produce violent racist content.[79]

We now outline how companies and other actors should respond to these risks in the near term. Our recommendations come in two buckets: (1) how to reduce the production and dissemination of harmful content; (2) how to empower users so that harmful impacts are reduced.

PROTECTING DEMOCRATIC INTEGRITY FROM AI: REDUCING THE PRODUCTION AND DISSEMINATION OF HARMFUL CONTENT

Current state of affairs

Many of the risks set out above are partly addressed by policies and practices adopted by major online content-distribution networks like the large social media platforms. These platforms generally already prohibit certain kinds of harmful misinformation, exclusionary speech (including hate speech, bullying, and harassment), nonconsensual intimate image sharing, and speech threatening or inciting violence.[80] Social media companies also commonly set out how their policies apply to AI-generated content (such as TikTok, which allows synthetic media of public figures only if it doesn't break any other content policies).[81]

For their part, developers of generative AI foundation models and public-facing user applications also already have prohibitions on some categories of content that users should not generate, and have restrictions and guardrails against usages which might cause harm.[82] Measures include refusing to provide information, such as electoral information,[83] or blocking specific prompts,[84] meaning that if a user requests certain kinds of content concerning a named person - such as a celebrity or politician - the tool may refuse to generate

76 Bond, S. February 2024. https://www.npr.org/2024/02/08/1229641751/ai-deepfakes-election-risks-lawmakers-tech-companies-artificial-intelligence; Yerushalmy, J. February 2024. https://www.theguardian.com/world/2024/feb/23/ai-deepfakes-come-of-age-as-billions-prepare-to-vote-in-a-bumper-year-of-elections#:~:text=AI%20deepfakes%20come%20of%20age,US%20elections%202024%20%7C%20The%20Guardian
77 Abilov, A. 2021. https://ojs.aaai.org/index.php/ICWSM/article/view/18113/17916; Leingang, R. February 2024. https://www.theguardian.com/us-news/2024/feb/10/social-media-ai-misinformation-election-2024#:~:text=As%20the%20United%20States'%20fractured,elections%20in%20the%20misinformation%20age.
78 Siegel, D. and Bennett Doty, M. February 2023. https://gnet-research.org/2023/02/17/weapons-of-mass-disruption-artificial-intelligence-and-the-production-of-extremist-propaganda/
79 Gilbert, D. February 2024. https://www.wired.com/story/gab-ai-chatbot-racist-holocaust/
80 E.g., https://transparency.fb.com/policies/community-standards; TikTok. https://www.tiktok.com/creators/creator-portal/en-us/community-guidelines-and-safety/community-guidelines/
81 TikTok. https://www.tiktok.com/community-guidelines/en/integrity-authenticity/
82 OpenAi. https://openai.com/policies/usage-policies Adobe. https://www.adobe.com/uk/legal/licenses-terms/adobe-gen-ai-user-guidelines.html
83 Hoskins, P. March 2024. https://www.bbc.co.uk/news/business-68551206
84 Midjourney. https://docs.midjourney.com/docs/community-guidelines

it. Ahead of the elections specifically, major tech companies have committed through the Munich Accord to collaborate on tackling deceptive AI electoral content through better risk assessment, developing more effective provenance and detection technology, and improving transparency.[85]

Improvements to be made

Although user rules and guardrails are already in place across the major AI models, these are not fool-proof. Rules setting out acceptable user behaviour are useful for norm-setting, deterrence, and potentially taking action against violators, but are currently insufficiently developed and defined.[86] There is nothing like the vast array of rules that we see concerning social media platforms' content moderation systems, yet the basic aim - harm prevention - is broadly the same. Indeed generative AI's current rules are much closer to the kind of crude, simplistic rules platforms had in the 2000s and early 2010s. In pursuing these changes, companies need to learn from the vast experiences of trust-and-safety teams within social media companies.[87]

While we recognise that the extensive rules developed by social media platforms may not be possible in the immediate short term for AI companies, even minimal clarifications would be an improvement on the status quo.

> **AP1.1:** Developers of generative AI foundation models and user applications should **develop and publish more specific policies concerning the content that users may and may not generate**, especially with respect to content that undermines democratic integrity. These policies should be the explicit basis of models' guardrails, creating a symmetry between expectations on users and on the company itself.[88]

Technical guardrails are often introduced retrospectively after vulnerabilities are found and exploited by bad actors. This has significant limitations. For example, widely shared deepfake images of Taylor Swift were revealed to have been created by a community on 4Chan who were (successfully) trying to get around existing blocks, by using different prompts or misspelling names.[89] *Post hoc* improvements do not alleviate the general risk of increasingly sophisticated prompt injection.[90] Although red-teaming is widely employed by AI developers, ensuring that these exercises are focused on specific democratic threats, as well as opening them up to greater public scrutiny,[91] would help further reduce risks.

Even with a conscientious effort by companies to improve guardrails, current technological limitations mean that *some* harmful content is bound to get through – especially due to malicious use. Moreover, open-source LLMs can still be used to develop generative AI tools that are publicly available with none of these guardrails; and tools can be explicitly developed which are designed and marketed as lacking substantial content guardrails.[92]

It should therefore be expected that some such content will find its way to the main forums of digital discourse, namely social media platforms, where it can be circulated and amplified.[93] This is why a multipronged approach to addressing threats from generative AI is needed. There is no one silver bullet solution.

85 Microsoft, February 2024. https://news.microsoft.com/2024/02/16/technology-industry-to-combat-deceptive-use-of-ai-in-2024-elections/; AI Elections Accord. 2024. https://www.aielectionsaccord.com/uploads/2024/02/A-Tech-Accord-to-Combat-Deceptive-Use-of-AI-in-2024-Elections.FINAL_.pdf

86 For example, OpenAI has rules against "generating or promoting disinformation [and] misinformation" but this is never defined; see OpenAI. https://openai.com/policies/usage-policies. This stands in stark contrast to the elaborate specifications of what kinds of misinformation is disallowed by the major social media platforms; e.g. Meta. https://transparency.fb.com/policies/community-standards/misinformation.

87 This is something the UCL Digital Speech Lab has already advocated. Digital Constitutionalist. https://digi-con.org/how-should-we-regulate-llm-chatbots-lessons-from-content-moderation/

88 E.g., OpenAI has begun this work: OpenAI. https://openai.com/blog/how-openai-is-approaching-2024-worldwide-elections

89 Lanxon, N. February 2024. https://www.bloomberg.com/news/articles/2024-02-05/taylor-swift-deepfakes-originated-from-ai-challenge-report-says; Ortiz, S. January 2024. https://www.zdnet.com/article/microsoft-adds-new-designer-protections-following-taylor-swift-deepfake-debacle/

90 Ortiz, S. January 204. https://www.zdnet.com/article/microsoft-adds-new-designer-protections-following-taylor-swift-deepfake-debacle/

91 Open AI, September 2023. https://openai.com/blog/red-teaming-network

92 UNESCO, 2023. https://unesdoc.unesco.org/ark:/48223/pf0000387483/PDF/387483eng.pdf.multi Pringle, E. November 2023. https://fortune.com/2023/11/06/grok-elon-musk-artificial-intelligence-bot-xai/; Morrish, L. March 2024. https://www.wired.com/story/dark-side-open-source-ai-image-generators/

93 Barrett, P.M. et al, February 2024. https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/65cfdd6c0b2733710e9e96b0/1708121452365/NYU+CBHR+Election+2024_Feb+16+UPDATED.pdf

AP1.2: Companies should **upscale "prompt-hacking" and "red-teaming" exercises ahead of elections to help identify and mitigate model misuse**, and should publish summaries of what they are doing in this regard.[94] Detailed versions which cannot be shared publicly without increasing safety risks should be made available to independent civil society organisations upon request (such as, in the UK, registered charities).

The emerging risk of generative AI constitutes a unique opportunity for platforms to revisit their content moderation practices. They should ensure that their policies are clear, comprehensive, and consistent,[95] including policies on harmful misinformation, intimate images, exclusionary speech, and violent content. They should also review advertising rules, including those governing political advertising, to ensure they are fit for purpose.[96] The key to success, in our view, lies not in creating new policies banning generative AI content, but rather in *doubling down* on the enforcement of rules against harmful content, regardless of whether it was generated by humans or machines. If a deepfake is objectionable because it conveys misinformation, it is this fact–rather than the technology that produced it–that is the reason for removing it. The harm of misinformation, after all, does not hinge on its provenance (which in any case is often very difficult to decipher). Existing platform policies that target generative AI content are largely confused.[97]

Social media policies also require rapid detection and enforcement to be effective. Measures such as breaking search functions in response to specific threats can be a useful firebreak, but are frequently implemented only after an incident has become widespread.[98] Platforms must also ensure that they have sufficiently well designed content moderation systems overall, as well as trained and supported human content moderators to ensure that their policies are being enforced accurately and transparently.[99] This will need to include literacy across a wide range of languages and political contexts to be protective in a multi-election year.

AP1.3: **Rather than create new *sui generis* rules for synthetic content, social media platforms should double down on the enforcement of rules against harmful speech and rules about advertising on their platforms for all users**, removing content that breaches their policies regardless of whether it is generated by human or machine.

The EU has taken a lead on this in publishing draft guidelines for consultation already under the DSA on protecting the integrity of elections.[100] In the UK, however, protecting democratic integrity online does not fall under the remit of a sole regulator, with Ofcom responsible for Online Safety (where the content is illegal or harmful to children), the ECHR responsible for equalities, the Electoral Commission responsible for regulating campaign finance, law enforcement responsibility (when conduct constitutes a criminal offence), and responsibility for AI regulation (once instituted) distributed amongst regulators. As such, there is a need for regulators, cross-regulatory bodies such as the Digital Regulation Cooperation Forum and relevant Government departments and law enforcement to work together to ensure that what obligations currently exist for these new technologies under the various different regulatory authorities - especially those with newer mandates - is clear both to companies and to the public.

94 With thanks to Kyle Taylor, pers. comm., 2024
95 Dad, N. February 2024. https://www.linkedin.com/posts/nighat-dad-3a937173_2024election-globalvote-activity-7164029505626447872-7Ugi?utm_source=share&utm_medium=member_desktop
96 Michael, C. November 2023. https://www.theguardian.com/technology/2023/nov/15/facebook-ads-2020-election-rigged-stolen-instagram-policy#:~:text=Biden%2C%20was%20stolen.-,Meta%20will%20now%20allow%20political%20advertisers%20to%20say%20past%20elections,or%20future%20elections%20are%20legitimate
97 Meta has faced criticism from the Oversight Board on their policies being 'incoherent'; Oversight Board 2023. https://www.oversightboard.com/decision/FB-GW8BY1Y3, citing public comments the UCL Digital Speech Lab submitted to the Board.
98 Reuters, January 2024. https://www.theguardian.com/music/2024/jan/28/taylor-swift-x-searches-blocked-fake-explicit-images
99 For a more general set of recommendations on fortifying the digital information ecosystem against electoral interference, see the report produced by Democracy Reporting International, Forum on Information & Democracy, and International Institute for Democracy and Electoral Assistance, entitled 'Protecting democratic elections through safeguarding information integrity' and available at: Democracy Reporting International, Forum on Information & Democracy, and International Institute for Democracy and Electoral Assistance, 2024. https://informationdemocracy.org/wp-content/uploads/2024/02/Protecting-Democratic-Elections-2024.pdf. ; Judson, E. 2021. https://eu.boell.org/en/2021/07/09/gendered-disinformation-6-reasons-why-liberal-democracies-need-respond-threat
100 European Commission, February 2024. https://digital-strategy.ec.europa.eu/en/news/commission-gathering-views-draft-dsa-guidelines-election-integrity

AP1.4: UK policymakers should **invest in independent research to monitor the volume, type and potential effects of synthetic content generated in the run-up to the current set of elections** as far as possible, as well as the risks it poses to the democratic principles of *Truth*, *Equality*, and *Non-violence*, with findings to inform the development of effective regulatory oversight further down the line *(see Action Plan 2.13)*.

AP1.5: Regulators and law enforcement should **assess the extent to which they could mitigate democratic risks acting within their existing mandates and issue specific guidance** clarifying how existing law and policy/regulation already applies to generative AI and the democratic risks we outlined here.[101] (See, for instance, the FCC in the US taking rapid action against AI-generated robocalls).[102]

PROMOTING DEMOCRATIC INTEGRITY WITH AI

Even if an individual piece of synthetic content seems harmless when taken on its own (such as a deepfake of a politicians doing something silly, or piece of LLM-generated text that contains a minor mistake about some policy issue) the growing prevalence of such content may cause citizens to doubt the accuracy of *any* content they encounter, or the reliability of *any* source of information. This is one of the primary risks of information disorder: the confusion over authenticity and provenance of information means citizens are not equipped to access reliable information.

This 'degradation of the information environment', as the UK Government refers to it[103] could lead to even more widespread scepticism and distrust, even of accurate information and reliable sources. Such distrust is readily exploitable by bad actors, including those wanting to deny actual events or promulgate conspiracy theories.[104] Moreover, it can undermine the kind of evidence-based deliberation and decision-making necessary to democracy and the confidence in democratic processes that is required for peaceful decision-making and transitions of power.[105] We must empower citizens to deal with a more confusing information environment during an electoral period.

Current state of affairs

Some of the large social media companies manage this broader kind of informational risk by prebunking misinformation,[106] or by verifying content (usually with the support of a third-party fact-checker). If content is found to be false or misleading, a warning label is appended, together with links to accurate information.[107] The offending content can also be de-amplified to reduce its reach. Although fact-checking alone cannot solve the harms of disinformation, it provides audiences with more context and resources to enable them to interact with and assess the information.

Meta recently announced their new labelling policy, which is illustrative of the direction in which all companies should move (some others, including TikTok and YouTube, have already done so).[108] Meta applies "Imagined with AI" labels to images created by its own AI tools. It also puts visible marks on its images, as well as invisible watermarks (and identifying metadata) within the image files themselves. Accordingly, when such content shows up on platforms, it can be (more) easily identified as synthetic. Meta reports that they are building tools to help them identify synthetic content produced by other companies' tools. It is now working

101 Yousif, N. February 2024. https://www.bbc.co.uk/news/world-us-canada-68240887
102 Ibid.; Farid, H. February 2024. https://contentauthenticity.org/blog/february-2024-this-month-in-generative-ai-election-season
103 DSIT, October 2023. https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper/frontier-ai-capabilities-and-risks-discussion-paper
104 Rini, 2021. "Weaponized Skepticism: An Analysis of Social Media Deception as Applied Political Epistemology," Political Epistemology (Oxford: Oxford University Press)
105 GMF, February 2024. https://securingdemocracy.gmfus.org/asd-publishes-ai-election-security-handbook/ .
106 Coulter, M. February 2024. https://www.reuters.com/technology/google-launch-anti-misinformation-campaign-ahead-eu-elections-2024-02-16/
107 Meta currently exempts politicians from fact-checking; some of us have argued elsewhere that this is a mistake: Fisher, S., Kira, B., Arabaghatta Basavaraj, K. and Howard, J. February 2024. https://doi.org/10.54501/jots.v2i2.170
108 TikTok, 2024. https://www.tiktok.com/community-guidelines/en/integrity-authenticity/#3; YouTube, 2024. https://support.google.com/youtube/answer/14328491?hl=en

to label synthetic content created by other companies that shows up on its platforms.[109] Disclosing realistic uses of AI in political advertising is also required.[110]

AI companies also currently often include information on the reliability of their tools, such as through the suggested activities they prompt, or through warnings they display.[111]

Improvements to be made

In enacting recommendation (6), content-distribution platforms should implement a standardised, blanket labelling regime for all synthetic content: most straightforwardly, content produced by their own in-house AI tools should be clearly marked as AI-generated. This is particularly important for realistic audio, visual, and multi-modal content, as this information will allow users to engage with it in a more informed and empowered way.

It will be difficult to enforce labelling of other synthetic content, as detection of AI-generated content is not highly accurate.[112] However, even without effective enforcement, it is valuable for users be to given tools to enable them to add labels to their own content regarding the provenance of content at the point of posting[113] - or flag content they see as potentially AI-generated.[114] This will help, we think, to promote healthy norms regarding synthetic content.

AP1.6: Content-distribution platforms should **require labelling synthetic user-generated content and ads**,[115] including by automatically labelling that which is produced by their own in-house tools, and by enabling users to label their posts or suggest the labelling of other posts.

The labelling effort should be assisted by developers of AI foundation models and applications,[116] who should implement watermarking protocols for the contents generated using their tools. In many cases watermarked synthetic media can then be identified by content-distribution platforms and labelled as AI-generated. We recognise that current watermarking is not completely effective and can be evaded, disguised or removed by those determined to do so, and this should be a priority for longer-term technological development *(see Action Plan 2)*.[117]

AP1.7: Developers of generative AI foundation models and user applications should **watermark the contents produced by their tools**, where it is feasible in the short term.[118] This is likely to predominantly be feasible for the largest companies, many of whom already have watermarks embedded in their tools.

Another important site of empowerment is within the interface of chatbot tools, which users might use to search for relevant information about political issues, candidates, and elections. Here it is vital that chatbot interfaces communicate their limitations–and their unsuitability for political research–to users.

109 Clegg, N. February 2024. https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/
110 Vanian, J., November 2023. https://www.cnbc.com/2023/11/28/meta-updates-political-ad-rules-to-cover-ai-generated-images-videos.html; Meta, January 2024. https://www.facebook.com/government-nonprofits/blog/political-ads-ai-disclosure-policy
111 Gemini Google, 2024. https://gemini.google.com/app
112 Bontcheva, K. February 2024. https://edmo.eu/wp-content/uploads/2023/12/Generative-AI-and-Disinformation_-White-Paper-v8.pdf
113 Alongside the platform-focused recommendations we put forward here, an earlier report by Demos, prepared in collaboration with Cavendish and entitled 'Generating democracy: AI and the coming revolution in political communications,' makes a complementary set of recommendations concerning political campaigners' approach to generative AI: https://demos.co.uk/wp-content/uploads/2024/01/Generating-Democracy-Report-1.pdf.
114 On a similar model to X's Community Notes feature. See X, 2024: https://help.twitter.com/en/using-x/community-notes
115 Vanian, J., November 2023. https://www.cnbc.com/2023/11/28/meta-updates-political-ad-rules-to-cover-ai-generated-images-videos.html
116 Partnership on AI, March 2024. https://partnershiponai.org/wp-content/uploads/2024/03/pai-synthetic-media-case-study-analysis-1.pdf
117 Elliott, V., February 2024. https://www.wired.com/story/meta-crack-down-ai-generated-fakes/ David, E., February 2024. https://www.theverge.com/2024/2/6/24063954/ai-watermarks-dalle3-openai-content-credentials; Pugalia, A., March 2024. https://www.techpolicy.press/the-race-to-detect-aigenerated-content-and-tackle-harms/
118 Content Authenticity Initiative, 2024. https://contentauthenticity.org/how-it-works

This campaign of user empowerment should be reinforced by trusted public bodies, civil society groups and journalists, who should identify reliable sources of information and encourage citizens to use these to check the accuracy of the content they encounter. Citizens should also be encouraged to seek out information proactively, rather than relying too heavily on information that may have been narrowly targeted toward them.

AP1.8: Developers of generative AI applications should **ensure their tools provide clear information to users about the potential inaccuracy of the content produced**, with an explanation that these tools are not reliable sources of factual information.

Users can also be empowered to protect themselves and others from potential AI-driven increases in online attacks during a politically volatile period.

There remains a risk that some harms still fall between policy cracks. Traditional strategies for dealing with political disinformation, such as labelling or fact-checking, are insufficient to combat the harms of emotive, harassing or sexualised disinformation. To manage acute risks more effectively, downstream interventions to flag harmful AI-generated content to platforms by those who are being harmed by it should be improved, beyond current reporting systems which are often slow, inconsistent and opaque.

AP1.9: Ahead of elections, social media companies should:

a. **Ensure there are transparent escalation systems and clear channels of communication** in place for those targeted to report harassment campaigns.[119]
b. **Ensure signposts are easily available to resources for further support for those targeted.**[120]

There are significant time pressures in achieving our recommendations comprehensively. However, one thing that could be very realistically achieved is for the political parties to start to set the norms and show leadership in protecting democratic integrity.

In the current absence of reliable automated labelling, it is crucial that democratic actors support these transparency efforts of their own accord. Political parties and candidates, in particular, should ensure that their own use of generative AI tools is clear to the public - not only disclosing when they are required to (e.g. in political ads), but taking active steps to demonstrate transparently how they are using these tools, and ensuring they do not contribute further to information disorder.

AP1.10: Political parties should **develop a cross-party compact on how generative AI is to be used transparently and ethically in election campaigning.** This should include a commitment to not amplifying content about any candidate or party that there are reasonable grounds to believe is materially deceptive.[121]

Finally, there is one step that can be taken by social media platforms in the short term that can help support the protection of democratic integrity.

To return to labelling, we recognise that it is not a panacea. The fact that content is AI-generated does not make it deceptive, and the fact it is not AI-generated does not make it reliable. It is worth investigating -

119 Judson, E. July 2021 https://eu.boell.org/en/2021/07/09/gendered-disinformation-6-reasons-why-liberal-democracies-need-respond-threat. UNESCO, November 2023. https://www.unesco.org/en/articles/technology-facilitated-gender-based-violence-times-generative-ai
120 Digital Rights Foundation. https://digitalrightsfoundation.pk/; Chayn. https://www.chayn.co/ Demos, Threats Women Journalists Face Online: Ana's Story and Samina's Story: https://www.youtube.com/channel/UCeO4Yd8qi4y4IsjrdYHhJQQ/videos; Digital Rights Foundation, 2022 https://digitalrightsfoundation.pk/wp-content/uploads/2023/07/Policy-Brief-Cyber-Harassment-Helpline-2022.pdf; Deelen & Voght, February 2024. https://www.irex.org/insight/what-media-and-civil-society-leaders-can-do-mitigate-technology-fueled-misogyny-2024
121 Demos, January 2024. https://demos.co.uk/wp-content/uploads/2024/01/Generating-Democracy-Report-1.pdf

especially given that many platforms are adopting labelling already - what effects transpire from labelling with regard to citizen trust.[122] There is a risk that inconsistent labelling practices could compound informational confusion.[123] However, we favour a provisional presumption in favour of labelling on the grounds that it increases the information available to citizens about the content they are consuming, with the proviso that ongoing research into labelling is necessary.

AP1.11: Social media platforms should **make data available to independent researchers** to support research into the effects of AI-generated content[124] and countermeasures such as labelling on the spread of deceptive content on their services during an electoral cycle.

122 See research suggesting potential negative externalities of labelling. Vasse'i, R.M. and Udoh, G. February 2024. https://foundation.mozilla.org/en/research/library/in-transparency-we-trust/research-report/#fitness-check
123 Ibid.
124 With thanks to Felix Simon, pers. comm., 2024

ACTION PLAN 2
PROTECTING AND SUSTAINING DEMOCRATIC INTEGRITY IN THE LONG TERM

In this section, we consider longer term risks to the ideals of democratic integrity, as well as the opportunities that generative AI could offer to democracy.[125] Since generative AI is emerging at a time when democratic resilience is low,[126] it is especially important to take the opportunity to shore up core democratic principles. We also consider what fundamental risks or blockers exist to generative AI upholding those principles, which will need to be addressed in order for the opportunities to be realised.

Action Plan 1 presented the minimum actions we should expect from stakeholders in response to urgent and acute risks this year. In Action Plan 2, we turn to what expectations we should have for stakeholders to deliver on democratic integrity over the longer term.

We do not repeat the recommendations in Action Plan 1 for reasons of brevity, but they are not, for the most part, one-off recommendations. We stress that our recommendations in Action Plan 1 should be undertaken on an ongoing basis or periodically, with some developments which we set out here. In this way, risks to democracy can continue to be mitigated effectively as technologies and politics co-evolve.

There are inherent limitations to the recommendations we set out in Action Plan 1. One, most notable, is an accountability deficit.[127] In the short-term, there is limited action that policymakers and regulators are able to take to ensure company adherence, due to the length of legislative processes, the need to draft codes of practice to implement even imminent or existing legislation, and limitations to regulators' existing mandates. There is therefore at least a need for civil society, the public and policymakers to call for companies to take these steps to demonstrate the appetite for future scrutiny.

125 See also: Jackson Schiff, K.and Schiff, D.S., November 2023. https://theconversation.com/generative-ai-like-chatgpt-could-help-boost-democracy-if-it-overcomes-key-hurdles-212664.
126 Ziblatt and Levitsky. 2018. How Democracies Die (Penguin).
127 With thanks to Kyle Taylor, pers. comm. 2024

There are also steps which require a greater level of technological development to be fully implemented: such as the integration of more sophisticated labelling or watermarking technologies

Finally, there is evidence to be gathered and lessons drawn which we simply do not know yet. What risks genuinely are the most prominent or harmful will become apparent through the election period itself.

As such, in addition to the recommendations above, we recommend, with a long-term lens:

AP2.12: AI companies and social media companies should **work together to deploy more interoperable watermarking solutions**, such as are being developed through the Content Authenticity Initiative technical standards, and more robust, 'maximally indelible' watermarks and disclosure of the AI-generation of content.[128,129]

AP2.13: UK policymakers, AI companies and social media companies should continue to **support independent research, through means such as funding and data access provision, to identify the lessons learnt from ongoing monitoring of synthetic content generated during the election period** and the risk it poses to the democratic principles of truth, equality and non-violence. These learnings can then be integrated into the deployment of future safeguards against democratic risk.

AP2.14: AI companies developing applications for text generation should explore how to **build reliable citations into their search results, enabling generative AI tools to provide links to reliable sources that can be independently checked** for any apparently factual information they produce, rather than only relying on content warnings about inaccuracy.[130]

AP2.15: UK policymakers, regulators, and civil society oversight bodies should **assess AI and social media companies for their efficacy post-election with regard to the actions recommended in Action Plan 1.** Such recommendations should also inform the regulatory duties and codes of practice that companies will be required to abide by and report against (e.g. codes of practice drafted by Ofcom in enforcing the Online Safety Act).

128 This recommendations builds on Action Plan 1 Recommendation 9

129 See the duties in the EU AI Act around transparency of outputs and disclosures of AI uses, as well as, https://contentauthenticity.org/; Evan Harris, D. and Norden, L. March 2024. https://spectrum.ieee.org/meta-ai-watermarks; see e.g. Halford, C., March 2024. https://www.bbc.co.uk/rd/blog/2024-03-c2pa-verification-news-journalism-credentials#:~:text=Like%20an%20audit%20trail%20or,where%20it%20has%20come%20from.

130 La Trobe University, 2024. https://latrobe.libguides.com/artificial-intelligence/referencing

CHALLENGES FOR EQUALITY

Generative AI models are trained on datasets which lead to those tools replicating and amplifying the biases of that data: and measures to try to combat this problem so far have been limited in effect. Challenges that must be overcome include:

CHALLENGE	EXAMPLES
Increased use of generative AI tools sees the biases in those tools being replicated and amplified,[131] entrenching stereotypes and biased narratives.[132]	Image generators producing disproportionately fewer images of women or of Black people,[133] or tending to portray white people or men in response to prompts of higher-paying jobs.[134]
Surface level fine tuning to address bias, can compound rather than challenge bias.[135]	Image generators producing 'diverse' images of historically - and relevantly - undiverse groups, such as Nazi soldiers.[136] Text generators refusing to take a stance on clear or settled ethical issues, or suggesting there is no right answer.[137]
'Astroturfing' (i.e., generating mass sham expressions of opinion) facilitated by AI crowds out space and reduces trust in constituent-representative interactions.	Legislators failing to distinguish between genuine and fake enquiries from constituents.[138,139]

OPPORTUNITIES FOR EQUALITY

In order for any opportunities to be realised, at a *minimum* the long-term risks of AI highlighted above would need to be addressed. Without a demonstration from AI companies and other relevant stakeholders that these risks can be adequately mitigated, pursuing opportunities will not be successful and could even be actively harmful.

However, generative AI may hold the promise of improving equality, by reinforcing equal regard, equal rights, and equal respect—and also by enabling more people to participate fully in the democratic process.

Specific opportunities for *Equality* include:

OPPORTUNITY	EXAMPLES
Synthetic content that empowers marginalised individuals or groups.	Large language models that provide contextualising historical facts concerning social inequalities. Large language models that engage in educational conversations without burdening members of that marginalised group to do so.[140]

131 IBM, October 2023. https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/ O'Neil, L., August 2023. https://www.rollingstone.com/culture/culture-features/women-warnings-ai-danger-risk-before-chatgpt-1234804367/
132 Demos, January 2024. https://demos.co.uk/wp-content/uploads/2024/01/Generating-Democracy-Report-1.pdf; UNESCO and IRC AI, 2024. https://unesdoc.unesco.org/ark:/48223/pf0000388971
133 Zhou, M., Abhishek, V. and Srinivasan, K. Date unknown. https://www.andrew.cmu.edu/user/ales/cib/bias_in_gen_ai.pdf
134 Nicoletti, L. and Bass, D. Date Unknown. https://www.bloomberg.com/graphics/2023-generative-ai-bias/ ; IBM, October 2023. https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/
135 This is not to say all fine-tuning is problematic - fine-tuning can be successful in some cases in reducing bias - see UNESCO and IRC AI, 2024. https://unesdoc.unesco.org/ark:/48223/pf0000388971
136 Powel, J. 2024. https://eu.usatoday.com/story/news/nation/2024/02/25/google-suspends-ai-image-feature-pictures-of-people/72737627007/
137 Kleinman, Z. February 2024. https://www.bbc.co.uk/news/technology-68412620
138 Kreps, S. and Kriner, D., March 2023. https://www.brookings.edu/articles/how-generative-ai-impacts-democratic-engagement/
139 European Parliament, October 2023. https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/751478/EPRS_BRI(2023)751478_EN.pdf; Kreps, S. 2023. https://www.whitehouse.gov/wp-content/uploads/2023/06/Kreps_PCAST.pdf
140 Although this may not be as effective as human counterspeech: see Benesch, S. and Buerger, C. March 2024. https://www.techpolicy.press/can-ai-rescue-democracy-nope-its-not-funny-enough/ Demos, November 2024. https://demos.co.uk/wp-content/uploads/2023/12/Drivers-of-Digital-Discord.pdf New Public Sphere, 2024. https://newpublicsphere.stir.ac.uk/; Bradley, E., January 2023. https://www.cbc.ca/news/canada/prince-edward-island/pei-black-history-month-evelyn-bradley-1.6725281

Generative AI tools that widen democratic participation.[141]	Large language models that facilitate written exposition and translation between languages, supporting communications among citizens, and with their democratic representatives.[142] Generative AI used by political candidates to better understand data about their constituents and constituencies.[143] Generative AI used to facilitate public deliberation.[144] Audio-visual tools that allow users to create high quality content without specialist skills. Civil society organisations representing marginalised groups are able to upscale through AI-enabled efficiencies.

CHALLENGES FOR TRUTH

Let us imagine that, in the future, AI-generated content comes to constitute a large proportion—perhaps even most—of the content we encounter (and, in turn, that generative AI models too increasingly ingest other synthetic content as input). We will want to ensure that this synthetic content is accurate, at least in contexts where users are expecting to receive facts. The content should also be relevant, informative, and clear, taking account of the user's needs.

CHALLENGE	EXAMPLES
Proliferation of low-quality AI content pervades information ecosystem.	AI-generated content driving up advertising revenue for junk news sites makes it harder for quality news to compete.[145] Increasing use of AI weakens rather than bolsters the news industry.[146] Increasing use of AI in content production risks increasing error and hallucination rates, and reducing audience trust.[147,148] Reduction in diversity of content.[149]

OPPORTUNITIES FOR TRUTH

Insofar as AI-generated content has the properties of being relevant, informative, and clear, taking account of the user's needs—and is widely recognised to have them—citizens can feel increasingly confident in the information environments they inhabit. This, we think, would help ground a public conversation in which people deliberate sincerely, with a shared understanding of basic facts and criteria for establishing the truth of claims.

141 European Parliament, October 2023. https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/751478/EPRS_BRI(2023)751478_EN.pdf; Jackson Schiff, K.and Schiff, D.S., November 2023. https://theconversation.com/generative-ai-like-chatgpt-could-help-boost-democracy-if-it-overcomes-key-hurdles-212664
142 Ibid
143 Purtill, J. February 2024. https://www.abc.net.au/news/science/2024-02-21/ai-elections-deepfakes-generative-campaign-endorsement-democracy/103483710
144 Collective Intelligence Project, 2024. https://cip.org/alignmentassemblies; Mowbray, A., February 2024. https://blogs.bath.ac.uk/iprblog/2024/02/22/how-ai-could-help-citizens-assemblies-make-well-informed-decisions/
145 Ryan-Mosley, T. June 2023. https://www.technologyreview.com/2023/06/26/1075504/junk-websites-filled-with-ai-generated-text-are-pulling-in-money-from-programmatic-ads/
146 Simon, F., February 2024. https://www.cjr.org/tow_center_reports/artificial-intelligence-in-the-news.php
147 Longoni, C. June 2022, https://dl.acm.org/doi/fullHtml/10.1145/3531146.3533077
148 Demos, January 2024. https://demos.co.uk/wp-content/uploads/2024/01/Generating-Democracy-Report-1.pdf
149 Samuel, S. April 2023. https://www.vox.com/future-perfect/23674696/chatgpt-ai-creativity-originality-homogenization

Specific opportunities for *Truth* include:

OPPORTUNITY	EXAMPLES
Accurate and helpful synthetic content about political candidates / parties, political processes and policies which enable more productive political discourse.	Personalised images or videos, showing a user how to fill out the specific ballot they will receive in their constituency race. Large language models that identify gaps, errors, and inconsistencies in the statements of political candidates or parties. Large language models that summarise legislative processes, laws, public opinion, or constituent feedback.[150] Large language models specially tailored to provide citizens advice services.[151,152] Independent and public interest news organisations are able to upscale delivery with the same resources, and enable more high-quality information to be produced and disseminated.[153,154,155,156,157,158]
AI-generated re-framings of targeted campaign materials.	Large language models that reproduce campaign materials to present the topics from different perspectives.

CHALLENGES TO NON-VIOLENCE

There are challenges relating to nonviolence which must be addressed over the medium to long term.

CHALLENGE	EXAMPLES
Models proliferate hate speech.	Chatbots producing content containing slurs or conspiracy theories.[159]
Violent and convincing content is easier to create and amplify, and more difficult to fact-check.	AI amplifies existing risks of provoking social unrest or inciting violence (especially in already inflammatory situations) as seen in earlier social media cases concerning the genocide against the Rohingya[160] and violence between communities in Leicester.[161] Environments in which atrocities can be perpetrated are worsened through increased disinformation.[162] Human counterspeech replaced with less effective AI-driven counterspeech.[163]

150 European Parliament, October 2023. https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/751478/EPRS_BRI(2023)751478_EN.pdf
151 See, for example, the Money Saving Expert ChatGPT, July 2023: https://www.moneysavingexpert.com/pressoffice/2023/ai-say--ai-say--ai-say--martin-lewis---mse-launch-revolutionary-/
152 Demos, January 2024. https://demos.co.uk/wp-content/uploads/2024/01/Generating-Democracy-Report-1.pdf
153 Ibid
154 Arguedas, A.R., and Simon, F. July 2023. https://www.oii.ox.ac.uk/wp-content/uploads/2023/08/BII_Report_Arguedas_Simon.pdf
155 Demos, January 2024. https://demos.co.uk/wp-content/uploads/2024/01/Generating-Democracy-Report-1.pdf
156 Beckett, C. September 2023. https://www.lse.ac.uk/News/Latest-news-from-LSE/2023/i-September-2023/Nearly-three-quarters-of-news-organisations-believe-generative-AI-presents-new-opportunities-for-journalism
157 AI Journalism Lab, 2024. https://www.journalism.cuny.edu/j-plus/ai-journalism-lab/
158 Talfan Davies, R., February 2024. https://www.bbc.com/mediacentre/articles/2024/update-generative-ai-and-ai-tools-bbc#:~:text=We%20set%20out%20that%20we,AI%20to%20support%20content%2Dmaking; Politico, March 2024. https://www.bundle.app/en/breakingNews/-i-would-not-blindly-trust-them---how-journalists-should-approach-ai-28C030AD-0D32-465E-8CC5-91BF3F81010A
159 Gold, A. May 2023. https://www.axios.com/2023/05/25/generative-ai-antisemitism-bias
160 Milmo, D. December 2021. https://www.theguardian.com/technology/2021/dec/06/rohingya-sue-facebook-myanmar-genocide-us-uk-legal-action-social-media-violence
161 Abdul, G. September 2022. https://www.theguardian.com/uk-news/2022/sep/19/mayor-blames-leicester-hindu-muslim-unrest-on-social-media-disinformation
162 Global Centre for the Responsibility to Protect, March 2024. https://reliefweb.int/report/world/relationship-between-digital-technologies-and-atrocity-prevention
163 Benesch, S. and Buerger, C., March 2024. https://www.techpolicy.press/can-ai-rescue-democracy-nope-its-not-funny-enough/

OPPORTUNITIES FOR NON-VIOLENCE

There could also be scope to support non-violence with the aid of generative AI, helping citizens to manage their disagreements through respectful engagement and voting instead of the use of force.

Specific opportunities for *Non-violence* include:

OPPORTUNITY	EXAMPLES
Synthetic content that encourages respectful engagement.	Large language models that suggest more reasonable / less combative rephrasing of threatening or uncivil speech.[164]

CURRENT STATE OF AFFAIRS

Currently, generative AI applications can be unreliable, with inaccuracies arising from poor quality data or "hallucinations" (whereby the AI generates novel falsehoods). They can also produce misleading content, due to biased input data or "sycophancy" (whereby the AI prioritises what the user seems to want over what is accurate). Although applications of generative AI technologies are proliferating rapidly, it is not clear that there has yet been much attention from policymakers on how they could be used to actively support democratic deliberation and decision-making.

The algorithms underpinning today's leading content-distribution platforms seem unlikely to systematically surface and boost the kinds of synthetic content that would support and strengthen democratic norms in public discourse. Instead, they tend to focus on user engagement, often promoting content that is attention-grabbingly uncivil or polarising, rather than making any constructive contribution to democratic discourse. There has been significant work into how platforms could be improved: but not yet widespread adoption from companies who still deploy engagement-maximising systems.[165]

Although there is widespread consensus on the kinds of general measures needed to improve AI, there is a patchwork of regulatory efforts around generative AI applications and the foundation models on which they are built. The UK government's emerging process guidance and AI assurance guidance includes recommendations on risk assessments,[166] monitoring for vulnerabilities, and auditing training datasets and monitoring biases arising from models.[167,168] The AI White Paper consultation response sets out the intention for further research, evidence-gathering and advice from specific regulators, but no regulation has yet come in.[169] The UK's Online Safety Bill has been passed, which will affect platforms' duties around illegal content and enforcing their terms of service, but what specific measures that will require of platforms is still under consultation.[170]

The EU has made significant strides in these directions through the EU AI Act, which has recently passed the European Parliament. The Act includes transparency requirements for providers of general-purpose AI models, of which large language models which power tools such as generative AI chatbots are a subset. These require transparency about technical documentation,[171] copyright adherence and publishing a summary about the training data used[172] (with some exceptions for open-source models).[173] Models which carry 'systemic risk' - which could include risks to democratic processes, democratic values and human rights, or

164 Stray, J. August 2023. https://knightcolumbia.org/content/the-algorithmic-management-of-polarization-and-violence-on-social-media; Stray, J., July 2022. https://arxiv.org/abs/2207.10192; Ovadya, A. May 2022. https://www.belfercenter.org/publication/bridging-based-ranking

165 Ovadya, A. May 2022. https://www.belfercenter.org/publication/bridging-based-ranking, Stray, J. July 2022. https://arxiv.org/abs/2207.10192, Stray, J. August 2023. https://knightcolumbia.org/content/the-algorithmic-management-of-polarization-and-violence-on-social-media, Center for Human Compatible Artificial Intelligence, January 2024. https://humancompatible.ai/news/2024/01/18/the-prosocial-ranking-challenge-60000-in-prizes-for-better-social-media-algorithms/, Thorburn, L. and Ovadya, A., October 2023. https://www.niemanlab.org/2023/10/social-media-algorithms-can-be-redesigned-to-bridge-divides-heres-how/

166 DSIT, October 2023. https://www.gov.uk/government/publications/emerging-processes-for-frontier-ai-safety/emerging-processes-for-frontier-ai-safety#responsible-capability-scaling

167 Ibid.; DSIT, February 2024. https://assets.publishing.service.gov.uk/media/65ccf508c96cf3000c6a37a1/Introduction_to_AI_Assurance.pdf

168 DSIT, October 2023 https://www.gov.uk/government/publications/emerging-processes-for-frontier-ai-safety/emerging-processes-for-frontier-ai-safety#executive-summary

169 DSIT, 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

170 Ofcom, 2024. https://www.ofcom.org.uk/online-safety/information-for-industry/roadmap-to-regulation

171 101, European Parliament, March 2024. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html

172 107, and Article 53. Ibid. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html

173 104, Ibid. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html

give rise to bias or discrimination,[174] also have more substantial risk assessment and mitigation duties.[175]

Where general-purpose AI systems are intended to be public-facing and used by the public - such as generative AI tools - they must be clearly described as an AI system.[176] Generative AI tools must also mark their outputs as 'artificially generated or manipulated', as far as is reasonably technically feasible.[177] Deployers of generative AI tools to create deepfakes, must also disclose that the content has been artificially generated or manipulated. Deployers of these tools to create text published 'to inform[] the public on matters of public interest' must do the same, unless there is human editorial oversight and responsibility in place.[178] These obligations will be supported by Codes of Practice on how they should be met.[179]

However, uncertainty remains at the EU level as well: there remain legislative steps to go through before the Act is officially law, and similar dependencies on Codes of Practice (as for the OSB) remain, with a significant time lag before rules come into effect and enforcement might begin.[180]

As encouraging as these developments are, we remain currently in a period of uncertainty, and the gap in democratic control leaves us with a market that is unlikely to support the development of public-interest AI tools.

PROTECTING DEMOCRATIC INTEGRITY FROM AI

Improving the quality of an AI-driven information environment

Developers of AI should orient their models and applications towards producing more accurate content in contexts where users are seeking factual information—including, especially, when they want to know about electoral processes, political entities, and policy issues. In doing so, developers will need to extend their efforts to understand and minimise AI-generated inaccuracies and biases—particularly given the risk that these feed further generative AI models, promulgating the harmful content in a vicious circle.

> **AP2.16:** Developers of AI models and applications should **put significant resources into understanding — and explaining — the provenance of AI-generated inaccuracies and biases, and take meaningful steps to rectify these** (e.g. through better curated training data, more human feedback, or more sophisticated guardrails).[181] This should broaden out from the scope of the Munich Accord[182] to cover harms to truth, equality and non-violence, and not only deceptive election content.

Turning to the content-distribution platforms, these would ideally move towards algorithmic designs that are both engaging—so that citizens want to use the platforms—and supportive of democratic principles—so that users can participate in civil discussion there. (It is worth noting that fostering such an environment will involve removing the all-too-likely onslaught of AI-generated spam and junk.)

> **AP2.17:** Content-distribution platforms should **conduct and publish risk assessments of the integration of generative AI tools into their services *before* they are integrated.** (Ideally, this would form part of their duties under the Online Safety Act.)

174 110, Ibid. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html
175 114-115, https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html
176 Article 50, 1. Ibid. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html
177 Article 50, 2. Ibid. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html
178 Article 50, 4. Ibid. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html
179 Article 50, 7. Ibid. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html
180 Dotan, R. 2024. https://www.linkedin.com/posts/ravit-dotan_timeline-activity-7173683281014452227-gwSc?utm_source=share&utm_medium=member_desktop; https://www.linkedin.com/feed/update/urn:li:activity:7173630965133561857/
181 E.g. Mitchell, M., February 2024. https://time.com/6836153/ethical-ai-google-gemini-debacle/
182 Tech Accord. https://securityconference.org/en/aielectionsaccord/

It is not only generative AI companies and social media companies who contribute to the information environment.[183] Other organisations in a wide range of industries are rapidly working to develop their understanding and guidance for how these tools should be used in their own contexts.[184]

News organisations in particular have made early strides in setting usage policies for their internal use of generative AI tools in creating content,[185] and have also begun to take a stand against their content being scraped by AI companies to train their AI tools,[186] which may create a market for quality information needed to produce more robust and reliable generative AI tools.[187]

As the generation of synthetic content is not yet widely regulated specifically, policy-makers should gather evidence from the current electoral cycle and feed this into proposals for future regulation of AI (including implementation of the EU's AI Act, as well as possible future legislation in other jurisdictions). This research would provide a foundation for effective oversight of AI companies' policies and guardrails in the medium term. This should also include monitoring of the efficacy of existing and new offences which cover synthetic content in reducing the incidence of harm.[188]

AP2.18: Industry standards should be **set within sectors to define sector-leading usage rules and best practice for generative AI tools**, which companies could then be certified on the basis of their compliance with.[189] These should be developed through collaboration between companies, regulators and civil society organisations.

AP2.19: UK policymakers should **impose obligations on AI companies requiring them to undertake comprehensive risk assessments**, with a focus on the risks that their models and products pose to democratic integrity. This should be enforced through meaningful audit by regulators and routes to data access for independent civil society organisations.[190] The newly-passed EU AI Act moves in this direction, with duties on developers of general-purpose AI models which meet the threshold for posing systemic risks (which can include models used for generative AI tools) to assess and mitigate these risks,[191] while the Digital Services Act provides for data access to social media platform data: but the UK is a way behind.

183 Ryan-Mosley, T. June 2023. https://www.technologyreview.com/2023/06/26/1075504/junk-websites-filled-with-ai-generated-text-are-pulling-in-money-from-programmatic-ads/

184 E.g. CampaignLab. https://www.campaignlab.uk/ ; Demos, June 2023. https://demos.co.uk/generative-ai-policy-paper/; Generative AI in the newsroom. https://generative-ai-newsroom.com/; Demos, January 2024. https://demos.co.uk/wp-content/uploads/2024/01/Generating-Democracy-Report-1.pdf

185 Maher, B., November 2023. https://pressgazette.co.uk/publishers/nationals/telegraph-generative-ai-guidelines-policy-copyright/

186 Milmo, D., September 2023. https://www.theguardian.com/technology/2023/sep/01/the-guardian-blocks-chatgpt-owner-openai-from-trawling-its-content#:~:text=The%20Guardian%20blocks%20ChatGPT%20owner,intelligence%20(AI)%20%7C%20The%20Guardian

187 Demos, December 2023. https://demos.co.uk/research/drivers-of-digital-discord-how-news-media-and-social-media-drive-online-discourse-and-pathways-for-change/; Irwin, L. February 2024. https://thehill.com/policy/technology/4492468-google-paying-independent-publishers-test-unreleased-generative-ai-platform/#:~:text=Google%20has%20announced%20it%20will,for%20receiving%20analytics%20and%20feedback

188 E.g. the UK criminalising the sharing of non-consensual intimate images a sexual offence whether real or 'made or altered by computer graphics or in any other way', and the EU Directive on violence against women criminalising non-consensual intimate image sharing, explicitly including the sharing of deepfakes, and the FCC banning robocalls. DSIT, January 2024. https://www.gov.uk/government/publications/online-safety-act-new-criminal-offences-circular/online-safety-act-new-criminal-offences-circular#introduction Milmo, D. October 2023. https://www.theguardian.com/technology/2023/oct/24/techscape-uk-online-safety-bill-clean-up-internet Ofcom, November 2023. https://www.ofcom.org.uk/__data/assets/pdf_file/0019/271243/volume-2-illegal-harms-consultation-1.pdf European Commission, March 2023. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0105 ; Equality Now, September 2023, https://audri.org/wp-content/uploads/2024/01/EN-AUDRi-Briefing-paper-deepfake-06.pdf

189 McNulty, L. February, 2024. https://www.cityam.com/the-notebook-the-city-has-to-unite-against-the-risks-of-generative-ai/; Demos, January 2024. https://demos.co.uk/research/generating-democracy-ai-and-the-coming-revolution-in-political-communications/

190 Albert, J. December 2022. https://algorithmwatch.org/en/dsa-data-access-explained/

191 European Parliament, March 2024. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html

PROMOTING DEMOCRATIC INTEGRITY WITH AI

As technical understanding improves, producers of these tools have a duty to ensure that their users can understand the ways in which they can best be used as well as their shortcomings.[192] This goes beyond simple product warnings. If products are marketed as enabling users to search for information, users may assume that results are surfaced in a similar way to a search engine. If conversations have the visual and linguistic appearance of talking with an agent, users may naturally infer that the 'ghost in the machine' is something like a human. A better approach would be to embed helpful signals in how tools and their outputs are described, presented, and marketed.

This extra clarity would mean that a greater diversity of generative AI tools which served different purposes would be safer for the public to engage with. In a world in which the risks of an incorrect answer are huge, generative AI tools must be limited in what answers they can reply - with guardrails put in so that they avoid offering strong views, or pronouncing on uncertain facts - which in some cases, can end up compounding rather than challenging harms.[193] With a more empowered public, however, who are better able to navigate using these tools, more space is opened up for generative AI tools with different purposes and limits and which communicate in different ways.

AP2.20: AI tool producers should **design the interfaces of their products and services to communicate effectively their purpose and limitations to users.**

Public Interest Generative AI

We would also like to see the development of democratically beneficial applications, such as large language models that present alternative perspectives or framings of arguments, or audio-visual tools that make complex information easier to understand.

AP2.21: Funding bodies in the public and private sector should **further incentivise the development of democratically beneficial generative AI applications**[194] (e.g. by supporting a 'Democratic Sandbox' for companies, civic tech and civil society organisations to collaborate in and experiment with developing public and open democratic AI systems; supporting the development of generative AI tools and best practice for public interest functions such as charities and public interest news organisations).[195]

AP2.22: AI companies should **publish the principles on which their AI tools have been designed and trained** — and how this has been achieved, with what oversight (e.g. through 'democracy-by-design' procedures, training procedures, and independent oversight or audit procedures).[196]

Some AI companies are attempting to tackle the more fundamental problem with bias and harmful outputs, namely by changing what AI models should be being trained to optimise for. Steering the direction of generative AI development in a more pro-social direction, to ensure the outputs are more likely to respect

192 Demos, December 2023. https://demos.co.uk/wp-content/uploads/2023/12/Drivers-of-Digital-Discord.pdf
193 Lin, B., February 2024. https://www.wsj.com/articles/google-and-anthropic-are-selling-generative-ai-to-businesses-even-as-they-address-its-shortcomings-ff90d83d ; BBC, February 2024. https://www.bbc.co.uk/news/technology-68412620
194 Tech Accord, February 2024. https://www.aielectionsaccord.com/uploads/2024/02/A-Tech-Accord-to-Combat-Deceptive-Use-of-AI-in-2024-Elections.FINAL_.pdf
195 Linklaters, February 2024. https://techinsights.linklaters.com/post/102izns/prepare-for-take-off-uks-digital-regulatory-cooperation-forums-ai-and-digital; Harvard University. https://huit.harvard.edu/ai-sandbox; DSIT, August 2023. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper; Taylor, K. February 2024. https://www.youtube.com/watch?v=MHgL9PZxits GMT, February 2024. https://www.gmfus.org/news/gmf-launches-project-pioneer-novel-technologies-strengthen-democratic-resilience; House of Lords, 2023 https://bills.parliament.uk/publications/53068/documents/4030; Ajder, H, July 2023. https://www.linkedin.com/posts/henryajder_a-new-national-purpose-ai-promises-a-world-leading-activity-7075041764121661440-bE6J/
196 DSIT, February 2024. https://assets.publishing.service.gov.uk/media/65ccf508c96cf3000c6a37a1/Introduction_to_AI_Assurance.pdf

democratic ideals, is itself a process which needs democratic input to make it more likely to succeed. We are supportive of projects already being undertaken by AI companies and civil society to investigate how to determine what values should drive AI development in a participatory and democratic way.[197,198] These projects are, however, voluntary, nascent, and the outcomes are in no way binding on AI companies to improve their practices or affect their development.

Ultimately, this will need to be supported by government action. There are many steps governments and regulators could take to support a vibrant information environment–from investing in media literacy to increasing support for sustainable public interest journalism.

AP2.23: Regulators should **collaborate to produce consistent guidance that can govern the development of industry best practice in use of generative AI.** UK regulators should set out this intention in their upcoming strategic guidance to be published at the end of April 2024.

AP2.24: UK policymakers should **engage with public deliberations on governance of generative AI** and support these to be scaled and implemented into policymaking processes.[199]

CONCLUSION

2024 is set to be a year of significant political and technological change, with tech companies scrambling to assure the public their products are safe (enough) to participate in - or control - the election information ecosystem. In the midst of a patchwork of different regulatory expectations globally, there are simple steps - imperfect, but necessary - which can help mitigate the acute risks to equality, truth and non-violence from synthetic content.[200] The longer-term vision, beyond this heightened political horizon, is for a future in which positive political and technological change mutually reinforce each other, through digitising democracy and democratising digital. Policymakers need to collaborate with the public to come up with answers to how values and tech can and should interact,[201] before the technology - and those who control it - come up with the answer for us.

197 Collective Intelligence Project, 2023. https://cip.org/research/democratizing-ai
198 Perrigo, B., February 2024. https://time.com/6684266/openai-democracy-artificial-intelligence/
199 Hono, S.Y., February 2024. https://openfuture.eu/blog/alignment-assembly-on-ai-and-the-commons/; Belgium24, February 2024. https://belgian-presidency.consilium.europa.eu/en/news/launch-of-citizens-panel-on-artificial-intelligence/
200 For more guidelines on protecting elections this year, see European Commission, February 2024. https://digital-strategy.ec.europa.eu/en/news/commission-gathering-views-draft-dsa-guidelines-election-integrity Democracy Reporting International, Forum on Information & Democracy, and International Institute for Democracy and Electoral Assistance, 2024. https://informationdemocracy.org/wp-content/uploads/2024/02/Protecting-Democratic-Elections-2024.pdf European Parliament, October 2023. https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/751478/EPRS_BRI(2023)751478_EN.pdf Conroy, M. February 2024. https://www.linkedin.com/posts/meghanconroy_assessing-ai-borne-risks-to-the-integrity-activity-7165799067711705088-mQ_w?utm_source=share&utm_medium=member_desktop
201 https://www.fastcompany.com/91022817/act-now-on-ai-before-its-too-late-says-unescos-ai-lead

Licence to publish

Demos – Licence to Publish

The work (as defined below) is provided under the terms of this licence ('licence'). The work is protected by copyright and/or other applicable law. Any use of the work other than as authorized under this licence is prohibited. By exercising any rights to the work provided here, you accept and agree to be bound by the terms of this licence. Demos grants you the rights contained here in consideration of your acceptance of such terms and conditions.

1 Definitions

a 'Collective Work' means a work, such as a periodical issue, anthology or encyclopedia, in which the Work in its entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole. A work that constitutes a Collective Work will not be considered a Derivative Work (as defined below) for the purposes of this Licence.

b 'Derivative Work' means a work based upon the Work or upon the Work and other pre-existing works, such as a musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Work may be recast, transformed, or adapted, except that a work that constitutes a Collective Work or a translation from English into another language will not be considered a Derivative Work for the purpose of this Licence.

c 'Licensor' means the individual or entity that offers the Work under the terms of this Licence.

d 'Original Author' means the individual or entity who created the Work.

e 'Work' means the copyrightable work of authorship offered under the terms of this Licence.

f 'You' means an individual or entity exercising rights under this Licence who has not previously violated the terms of this Licence with respect to the Work, or who has received express permission from Demos to exercise rights under this Licence despite a previous violation.

2 Fair Use Rights

Nothing in this licence is intended to reduce, limit, or restrict any rights arising from fair use, first sale or other limitations on the exclusive rights of the copyright owner under copyright law or other applicable laws.

3 Licence Grant

Subject to the terms and conditions of this Licence, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) licence to exercise the rights in the Work as stated below:

a to reproduce the Work, to incorporate the Work into one or more Collective Works, and to reproduce the Work as incorporated in the Collective Works;

b to distribute copies or phono-records of, display publicly, perform publicly, and perform publicly by means of a digital audio transmission the Work including as incorporated in Collective Works; The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. All rights not expressly granted by Licensor are hereby reserved.

4 Restrictions

The licence granted in Section 3 above is expressly made subject to and limited by the following restrictions:

a You may distribute, publicly display, publicly perform, or publicly digitally perform the Work only under the terms of this Licence, and You must include a copy of, or the Uniform Resource Identifier for, this Licence with every copy or phono-record of the Work You distribute, publicly display, publicly perform, or publicly digitally perform. You may not offer or impose any terms on the Work that alter or restrict the terms of this Licence or the recipients' exercise of the rights granted hereunder. You may not sublicence the Work. You must keep intact all notices that refer to this Licence and to the disclaimer of warranties. You may not distribute, publicly display, publicly perform, or publicly digitally perform the Work with any technological measures that control access or use of the Work in a manner inconsistent with the terms of this Licence Agreement. The above applies to the Work as incorporated in a Collective Work, but this does not require the Collective Work apart from the Work itself to be made subject to the terms of this Licence. If You create a Collective Work, upon notice from any Licensor You must, to the extent practicable, remove from the Collective Work any reference to such Licensor or the Original Author, as requested.

b You may not exercise any of the rights granted to You in Section 3 above in any manner that is primarily intended

for or directed toward commercial advantage or private monetary compensation. The exchange of the Work for other copyrighted works by means of digital file sharing or otherwise shall not be considered to be intended for or directed toward commercial advantage or private monetary compensation, provided there is no payment of any monetary compensation in connection with the exchange of copyrighted works.

c If you distribute, publicly display, publicly perform, or publicly digitally perform the Work or any Collective Works, you must keep intact all copyright notices for the Work and give the Original Author credit reasonable to the medium or means You are utilizing by conveying the name (or pseudonym if applicable) of the Original Author if supplied; the title of the Work if supplied. Such credit may be implemented in any reasonable manner; provided, however, that in the case of a Collective Work, at a minimum such credit will appear where any other comparable authorship credit appears and in a manner at least as prominent as such other comparable authorship credit.

5 Representations, Warranties and Disclaimer

a By offering the Work for public release under this Licence, Licensor represents and warrants that, to the best of Licensor's knowledge after reasonable inquiry:

i Licensor has secured all rights in the Work necessary to grant the licence rights hereunder and to permit the lawful exercise of the rights granted hereunder without You having any obligation to pay any royalties, compulsory licence fees, residuals or any other payments;

ii The Work does not infringe the copyright, trademark, publicity rights, common law rights or any other right of any third party or constitute defamation, invasion of privacy or other tortious injury to any third party.

b Except as expressly stated in this licence or otherwise agreed in writing or required by applicable law, the work is licenced on an 'as is' basis, without warranties of any kind, either express or implied including, without limitation, any warranties regarding the contents or accuracy of the work.

6 Limitation on Liability

Except to the extent required by applicable law, and except for damages arising from liability to a third party resulting from breach of the warranties in section 5, in no event will licensor be liable to you on any legal theory for any special, incidental, consequential, punitive or exemplary damages arising out of this licence or the use of the work, even if licensor has been advised of the possibility of such damages.

7 Termination

a This Licence and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this Licence. Individuals or entities who have received Collective Works from You under this Licence, however, will not have their licences terminated provided such individuals or entities remain in full compliance with those licences. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this Licence.

b Subject to the above terms and conditions, the licence granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different licence terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this Licence (or any other licence that has been, or is required to be, granted under the terms of this Licence), and this Licence will continue in full force and effect unless terminated as stated above.

8 Miscellaneous

a Each time You distribute or publicly digitally perform the Work or a Collective Work, Demos offers to the recipient a licence to the Work on the same terms and conditions as the licence granted to You under this Licence.

b If any provision of this Licence is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this Licence, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

c No term or provision of this Licence shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

d This Licence constitutes the entire agreement between the parties with respect to the Work licenced here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This Licence may not be modified without the mutual written agreement of Demos and You.

DEMOS

Demos is a champion of people, ideas and democracy. We bring people together. We bridge divides. We listen and we understand. We are practical about the problems we face, but endlessly optimistic and ambitious about our capacity, together, to overcome them.

At a crossroads in Britain's history, we need ideas for renewal, reconnection and the restoration of hope. Challenges from populism to climate change remain unsolved, and a technological revolution dawns, but the centre of politics has been intellectually paralysed. Demos will change that. We can counter the impossible promises of the political extremes, and challenge despair – by bringing to life an aspirational narrative about the future of Britain that is rooted in the hopes and ambitions of people from across our country.

Demos is an independent, educational charity, registered in England and Wales. (Charity Registration no. 1042046)

Find out more at **www.demos.co.uk**

DEMOS

PUBLISHED BY DEMOS MARCH 2024

15 WHITEHALL, LONDON, SW1A 2DD
T: 020 3878 3955
HELLO@DEMOS.CO.UK
WWW.DEMOS.CO.UK