

[REDACTED]

From: FOI
Sent: 08 March 2024 09:51
To: [REDACTED]
Subject: FOI 008-24 - Response
Attachments: Oxford GenAI Summit notes.pdf; SMC policies.pdf; AI Safety Summit - Notes.pdf

Dear [REDACTED],

Our Ref: FOI-008-24

Thank you for your email to the Electoral Commission dated 18 January 2024.

The Commission aims to respond to requests for information promptly and has done so within the statutory timeframe.

Your request is shown below followed by our response.

Under FOIA please provide documents held by the electoral commission and produced after 1/1/2022 which reference the potential or actual impact of artificial intelligence upon the areas regulated by the commission.

If for reasons of time you need to further limit the scope of this request you may confine your search to documents of which staff at director level or above have had sight of.

Our response is as follows:

We hold the information you have requested.

The information we hold includes notes from several conferences that Electoral Commission staff have attended, and a paper summarising the policies of social media companies around political campaigning, including with regard to generative AI. Please see the attached documents we are disclosing to you.

Section 36 of the Freedom of Information (FOI) Act 2000

Some of the information that you have requested is exempt from disclosure under section 36(2)(b)(i) and (ii) and section 36(2)(c) of the FOI Act. This includes papers prepared to support discussions by the Electoral Commission Board at scheduled meetings; papers and notes relating to a roundtable of UK regulators; briefings to prepare a member of the Executive Team for media interviews; notes of meetings with social media companies; and internal briefing documents and discussion papers.

Section 36(2)(b) provides an exemption where disclosures would or would be likely (in the reasonable opinion of the qualified person) to inhibit:

- (i) the free and frank provision of advice or
- (ii) the free and frank exchange of views for the purposes of deliberation.

Section 36(2)(c) provides an exemption where disclosures would otherwise prejudice, or would be likely otherwise to prejudice, the effective conduct of public affairs.

Section 36 is a qualified exemption which requires the Public Interest Test (PIT) to be considered.

One of the Commission's functions is providing transparency about the election process. As an organisation, we are committed to making data available to voters, administrators and government. Refusing a request for disclosure runs contrary to those principles and functions.

On the other hand, Commissioners, senior staff and other regulators need a safe space (free from outside scrutiny and interference) to consider the broad implications of future technologies such as artificial intelligence before determining appropriate actions for the Commission to adopt in response.

Releasing background briefing documents prepared to support such broad discussions would be likely to inhibit the free and frank exchange of views for the purposes of deliberation in future. Commissioners and senior staff may come under political or media pressure to adopt particular views, and may be less willing to participate in discussions or explore challenging options.

Releasing detailed notes of meetings with other regulators or social media companies would be likely to inhibit the free and frank provision of advice and would be likely to prejudice the effective conduct of public affairs in future. Representatives of other organisations or private companies would be less likely to share relevant information with the Commission if they considered that the information could be released in response to future FOIA requests.

After carefully considering the opposing public interest arguments in favour of engaging the exemption and in favour of releasing the information, there is, we believe, a strong and overriding public interest in applying the exemption.

Section 40(2) of the Freedom of Information Act 2000

You will also notice that certain personal details have been redacted from the documents we are releasing. Section 40(2) of the FOI Act provides for an exemption where the information requested constitutes personal data as defined by the Data Protection Act 2018 (DPA), and where release of the information requested would breach one of the data protection principles.

Some of the information contained in the requested information falls within the description of personal data as defined by the DPA because the information relates directly to an identifiable living individual. This includes names and contact details of individuals in some instances. The individuals in some cases are junior staff members and they would not reasonably expect their information to be released.

The Commission strives to be an open, transparent authority, but in some circumstances we cannot responsibly release requested information, and we ask for your understanding in this regard.

If you are not satisfied with this response, please note that the Commission operates a review procedure, details of which can be found on the Commission website at:

<https://www.electoralcommission.org.uk/freedom-information/make-a-freedom-information-request>.

Please also note that if you have exhausted all internal Commission review procedures and you are still not satisfied you have the right to appeal to the Information Commissioner. Details of this procedure can be found on the ICO website: <https://ico.org.uk/>.

Yours sincerely

Information Officer

FOI@electoralcommission.org.uk

The Electoral Commission

electoralcommission.org.uk

Oxford GenAI Summit notes

21-22 October

Attended by [REDACTED] and [REDACTED] – please reach out if you have any questions. A detailed agenda of the conference can be found [here](#), while the conference website is accessible [here](#).

Contents

Key takeaways	2
Talks about democracy and elections	3
A race to regulate AI?	3
Digital content provenance: the foundation for a more authentic and transparent internet	3
Deep learning, deep fakes and the fight for democracy in the era of AUI.....	4
Other talks	6
Fact and fiction in the age of generative AI	6
How AI thinks	6
Generative AI, privacy, and data protection.....	7
Multi-stakeholder responses to deepfakes and synthetic media	8
AI safety is a social problem.....	9
AI governance and global politics.....	10
Generative AI and the future of journalism	11
Generative AI, disinformation and extremism.....	12

Key takeaways

- As expected, there was a large amount of ‘AI optimism’ at the conference, with the vast majority of attendees convinced that generative AI was going to cause significant changes to vast swathes of our lives. However, few people felt that apocalyptic outcomes were likely (see Nigel Toon’s talk for more) given current technology. In her address, Rachel Coldicutt said it would be more useful to look at apocalyptic scenarios and more at the ‘here and now’ and address risk posed currently.
- Almost all of the talks made some mention of the effects of generative AI on democracy and elections. There is space for the Commission to take the lead in this area, as the Commission’s own steering group on AI has discussed.
- There was also widespread agreement that generative AI detection software would not work effectively enough. Instead, transparency, provenance and education were held up as solutions to the proliferation of AI-generated content (*although I’m sceptical that any of these are sufficient*).
- Many of the discussions spoke about the need for collaboration – between regulators, governments (internationally) and private companies. Some speakers (Marco Mendicino) gave recommendations for policies which could be enacted to protect elections, while others (Chloe Smith) asserted existing laws within RPA and PPERA could be used for this purpose.
- There could be further opportunities for the Commission to further its use of AI technology itself – for example using machine learning to scrape information from different websites about candidates to create better public information for voters.
- There were a large number of people, organisations, or groups we should be reaching out to. These include: Truepic; Adobe; Microsoft; Marco Mendicino MP (Canadian House of Commons); the Coalition for Content Provenance and Authenticity; Newguard; and Meta.
- There were several individuals on panels and that we spoke to that we could reach out to about running a Learning Hour at the Commission.

Talks about democracy and elections

A race to regulate AI?

Chloe Smith MP

- Government policy is to act before mistakes happen. The frontier AI taskforce is working similarly to the Covid vaccine taskforce, partnering with other organisations, who provide government with access to their models.
- Hope is that global partner collaboration will lead to a rotating/annual discussion of the issues.
- When thinking about regulation, important to remember that frontier AI is just one part of AI. Having a central risk function is seen as key to understanding and responding to the whole AI ecosystem.
- On electoral registration, CS talked about how offences (*she clarified to us afterwards that she was thinking about Sections 106 and 115 of RPA*) will change with AI allowing material to be produced more quickly and in greater quantities. Against this she contrasted the ability of regulators to act, both in terms of enforcement actions and in terms of partnerships (especially with the police).
- Office for AI is producing a risk function, in collaboration with other regulators.
- Details and exemptions are among the hardest parts of any bill, particularly an AI bill. Hence the EU's difficulty with the AI Act, and why the UK government is proceeding as it currently is.

Digital content provenance: the foundation for a more authentic and transparent internet

Mounir Ibrahim, Trupic and Andrew Jenks, Microsoft

- This panel was all about transparency about the origin of content.
- Truepic and Microsoft are collaborating to capture cultural destructions in Ukraine, providing provenance details of images. The platform lets users authenticate images and detect changes to these images (C2PA Open Standard).
- These standards can be applied to synthetic material at the point of creation – shows whether content is synthetic or not.
- AJ: transparency is the best practice, rather than banning synthetic media.
- A 'Content Authenticity Initiative' was mentioned – *worth investigating*.
- Synthetic vs authentic is a false dichotomy, they have 14 steps on the C2PA scale in distinguishing between the two.
- Trust (re. information) and transparency are different topics.

- Q about media: chyron useful for news programmes to establish authenticity, as is embedding provenance into everything.
- Re-use of existing stock imagery is the biggest cause of misinformation.
- In terms of flagging text as AI-generated, they didn't think you can produce something that travels with the text. It's easier if the text is contained within a document or pdf.
- Microsoft have a 'Democracy Forward' team – *worth investigating*.

Deep learning, deep fakes and the fight for democracy in the era of AI

Marco Mendicino, MP in Canadian Parliament

Democratic institutions are looking antiquated in relation to AI and the challenges it poses.

Set the scene, in 2022 there were 6.2bn smartphone users

More people on social media, which has become a 'megaphone for disinformation' – people are getting 'bullied out of the digital town square'

Pew research – [few people feel they can influence the political system](#)

- Some pros to use of AI in elections include; using augmented realities to depict, for example, what the town centre will look like once regenerated, using ChatGPT to scrape data from across sites to compare candidates in an election (*I did think is there something there for modernising?*), publishing manifestos in different languages and communicating with voters in different languages; law enforcement using AI to foil cyberattacks
- Some cons include disinformation, deepfakes of politicians doing/saying things they didn't say/do, eroding trust, confidence, undermining result of elections via election-related disinformation

He has done research of principles of governance around AI and has come up with the following list – called 'Generally accepted AI principles' or GAIP

I think a lot of these align with our own remit – which I think could provide a useful hook for some EC policies around these areas.

1. Transparency
2. Equity
3. Safety
4. Privacy
5. Accountability
6. Responsibility and the public good

7. Openness and digital literacy

He then presented some policy idea/action based around these principles

- It's impractical to ban AI in political campaigns, so the question is about what safety produces can be put in place. Government deployment of AI should rest on public safety: with its use authorised by a law that balances rights (including privacy), with its use being 'reasonable', and overseen by the courts
- Create an offence for disinformation about the voting process – the Canadian Elections Commissioner has recently advocated this change.
- Create an offence around deepfakes depicting candidates breaking the law or committing violence – *I am unclear on who this offence would be targeted at – promoters, creators, distributors – but I wonder if there are existing offences in the RPA that could be used for this purpose?*
- Create an international convention based around the GAIP principles, build consensus across jurisdictions, work cross-institutionally and with concrete timelines – I agree this would be useful given the cross-jurisdictional nature of many of these policies – I think using IFES Community of Practice into Online Campaigning could be a useful starting point
- Campaigners should have responsibility to reduce harm – *perhaps something there around a code of conduct for campaigners, around use of AI and disinfo?*
- Watermarking
- Higher transparency on targeting/algorithms
- Model transparency
- Public education
- MM recommended using the G7 Rapid Response Mechanism and the OECD Global Incident Monitoring Process to track and report AI-generated disinformation relating to elections.
- Human oversight is needed in elections for accountability to be present. MM spoke about the recommendations of Canada's privacy commissioner in this area.

He also said a gap analysis was needed on how AI impacts elections

I spoke to Marco after and he seemed keen to feed in and work with us in the future.

Other talks

Fact and fiction in the age of generative AI

Sir Nigel Shadbolt

- Models need to scale in order to reach higher levels of competence, which requires compute and data. Chat GPT4 allegedly has 1 trillion parameters – but are we running out of data? The latest Google models (PaLM2) use less data, but still produce high quality – this requires less compute (also better for the environment).
- Benchmarking (measuring the quality of models) is an industry in itself.
- Generative AI is smart, but not smart like us – these applications don't have a mind of their own.
- IP claims from content generators have the potential to be hugely consequential.
- Data is key: what do we use? How do we use it?
- How do we align generative AI with broad human values, rather than narrow corporate ones?

How AI thinks

Nigel Toon (Graphcore)

- AI is going to augment personal intelligence, not replace or surpass it.
- We tend to overestimate our conscious intelligence and the amount of compute required, against that required for subconscious intelligence (e.g., hitting a tennis ball).
- AI is made from three ingredients:
 1. Information
 2. Semi-conductors – we are reaching physical limits to processing power here. Quantum computing will help somewhat, but it is not a full computer by itself (e.g., regarding memory)
 3. Learning methods – long way to go here
- ChatGPT is just a minimum viable product, it's not that impressive.
- NT distinguishes between points solutions (e.g., Google Maps) vs system solutions (e.g., Uber, which uses Google Maps to construct a whole business ecosystem). At the moment, we're primary on point relations regarding AI.

- Generative AI is really quite simple – it's like very good text prediction. It surprises us because it acts like a human. We'll see larger, multi-modal models able to solve more complex problems.
- On safety issues, it's important not to blame the machine. It has been put together by a human, with the problems already known. These include: bias, privacy issues, identity, persuasion, weapons, and ethics.
- Believes risk of singularity is overstated (heavily based on fiction) especially with silicon-designed chips.

Generative AI, privacy, and data protection

Karishma Brahmhatt (Allen & Overy); Camilla Winlo (Gemserv); Rafaela Nicolazzi (Google); chaired by Steve Wood (Allen & Overy)

On challenges and risks:

- KB: patchwork of laws in UK around AI. Other factors affecting risk include: data being used; societal impact; security measures in place (e.g., against data leakage/in response to automated decision-making).
- KB: ICOs of the world will look at data minimisation, purpose limitation (e.g., secondary uses not intended when data originally collected), and accuracy, when looking at enforcement.
- CW: four types of company re. AI:
 - 1) Ones who don't think they're using AI, but are concerned about its impact. They may have banned AI, but untrained employees will be using it on their devices, creating data risks.
 - 2) More confident organisations, who are building an AI strategy.
 - 3) Firms beginning to think about using 3rd party software – this introduces data/security risks. Building of software depends on the jurisdiction (e.g., scraping of data – if data removed due to GDPR, generative AI will change).
 - 4) Companies producing their own models (using client-specific or general datasets).
- RN: Google seeking more clarity from ICO about how privacy laws apply to generative AI. Company is working on policy recommendations in this area.
- RN: Google wants: dialogue; interoperability; sandboxes; and innovation/research.
- RN: Praised DRCF as an example of interoperability.

On governance:

- KB: need to ensure everyone is speaking the same language; that framework systems exist (also key for due diligence); data minimisation; and supply chain management (includes due diligence).

- CW: need senior team buy-in – difficult because this is a new topic.
- RN: Bard aims to promote data protection, as it includes a data privacy hub.

Multi-stakeholder responses to deepfakes and synthetic media

Claire Leibowicz (Partnership on AI); Laura Herman (Adobe); Bruce MacCormack (Project Origin & Coalition for Content Provenance and Authenticity); Josh Benaron (Irys); chaired by Mounir Ibrahim (Truepic)

On stakeholders:

- CL: intentional decisions are already being taken about what generative AI systems can produce (e.g., refusal to produce election disinformation). In February, the Partnership on AI launched a framework to delineate between builders, creators, and distributors of synthetic material.
- LH: Adobe (have generative AI application called Adobe Firefly) considers there to be three groups: creators, consumers, and implementers.
- LH: recent work integrating content recognition with Photoshop. Different groups (see above) have different interests, even creators are multi-stakeholder.
- BC: what happens if someone fakes a channel, rather than just the content? It's clear that detection isn't going to work. Rather, provenance is key. This requires buy-in from wide range of actors: news teams; camera operators; Adobe; etc.
- BC: misinformation is a tragedy of the commons, and requires a community response.
- JB: big blocker to all of this is social consensus. Is blockchain a way to help guarantee authenticity? ████████: *no*.

Aligning incentives between different stakeholders:

- BC: these incentives don't have to align, as differences can be strengths. Important to find the intersections between these through conversation.
- LH: research your different stakeholders and find highest common priorities.
- CL: uncertainty about issues (e.g., reporting on war) can lead to questions about wider topics like truth and democracy. There is beauty in the co-creation of AI, where humility can help a great deal.

Role of government

- BC: government increase in knowledge is slower than the rate of innovation. The knowledge within government comes from user experience, which ignores the development of infrastructure or processes which are out of sight. However, the government should be a proxy for society in its ability to consider and defend the public good.

- LH: Adobe have done research into providing the content creator's identity. People are often uncertain why the information is there, and generally just want to know whether they can trust what is in front of them. Adobe wants to avoid this, but it might be up to news organisations and social media companies to provide this additional information.
- CL: government is made up of different agencies with their own different interests, and there are trade-offs to be made between issues such as transparency and privacy. However, the alternative to regulation is large companies setting the rules – there is genuine appetite for regulation from OpenAI and others.
- CL: watermarking breaks down with text production.

On workers' rights:

- CL: prioritisation of stakeholders is key, and workers are certainly a central part of this. However, workers not a monolithic group.
- BC: deepfakes/synthetic material going to affect us all, not just workers.

On a world where people have different values:

- LH: people will grow to question sources where they don't see certain verification signs (e.g., blue tick on Twitter back in the day). Over time, they can be brought back into 'our' world.
- BC: people have critical skills too – technology is just a transmission vehicle.
- CL: doing some work on how easy it is to remove watermarks. More research is required on the effect of synthetic material and on companies' responsive methods.

AI safety is a social problem

Rachel Coldicutt (Careful Industries & Promising Trouble)

- 31/10/23 – AI and Society Forum
- We should focus on AI risk now, not future existential threat, to act before issues get too big to solve .
- OSB has seen a lot of specific risks included, rather than holistic consideration of threats (e.g., the specific offence of upskirting vs something about appropriate photography). A risk-based approach is technocratic, but we can't hope to capture all of them/or in their entirety.
- Instead, a rights-based approach speaks to the complexity of real life.
- Users should be central in any consideration of networks and risks. Most technologies are unfinished until used by people.
- Risk-based approaches assume no evolution of technology, or changes in the way it is used by users.

- Four key questions:
- 1. Should this exist?
- 2. How will others adapt it?
- 3. How will it be maintained?
- 4. How will we close it down?

On whether some uses of AI should be banned:

- Beyond privacy concerns, the rule of law does already exist, limiting what AI applications should be able to do to some extent.

On defining rights:

- Rights are already defined at the international and domestic level. It is up to AI companies to comply with these, as it's not the job of rights to shift to accommodate AI.
- There are no rights if we are not a community. The move to individual rights reduces collective agency.

On historical precedents:

- It's easy for people in this field to try to come up with new solutions, rather than drawing on previous work (such as on proposals for social media regulation).

AI governance and global politics

Carl Miller (Intelligence Squared, formerly with Demos); Kir Nuthi (Startup Coalition); and Keegan McBride (University of Oxford); chaired by Will Shao (University of Oxford)

- KM: Infrastructure surrounding and under AI is the most important thing, as part of the broader physical transformation.
- CM: choice of metaphors will guide future foreign policy action.
- KN: did project with Tony Blair Institute and Onward. Findings were that it is imperative to nail the fundamentals: access to capital, compute data, open data; hiring the best people; and navigating the regulatory environment.
- KM: three factors will determine who wins AI race: control of AI infrastructure; strategic integration into private sector; and the establishment of a regulatory system that enables this integration.
- KM: argues that democratic AI is open-source AI – *although I would dispute this*.
- CM: open-sourcing may prevent application of regulation to AI. As a technique, it just trades one form of capitalism for another.

- In a debate about the UK's international position regarding AI, there was consensus that it was in a strong position on safety, but miles behind the US on infrastructure, capital and compute.
- CM: AI creates new threat vectors that go beyond deepfakes (which already existed). It offers the opportunity to manipulate people in a more effective way. One example might be bots creating significant engagement with voters on social media platforms.
- CM: UK has a very different regulatory culture than the US, and should aim for control rather than laissez-faire.
- CM: sees US-led and Chinese-led of internet in the future. Thinks there is insufficient media attention on these changes and divisions, which could lead to greater authoritarianism in internet policy.

Generative AI and the future of journalism

Veena McCooles (NewsGuard); Felix Simon (University of Oxford); and Charlotte Jee (MIT Technology Review); chaired by Kenneth Cukier (The Economist)

- CJ: we're not getting AI to write articles – there are too many hallucinations. It might be used to produce headlines, transcripts, and summaries of data.
- VM: unreliable AI-generated news (UAIN) are websites specifically created for low amounts of content and high amounts of ads. NewsGuard have identified over 500 of these AI-generated sites with no editorial oversight, no doubt there are more.
- VM: when defining whether a site is UAIN, consider: is the majority of the content AI-generated? Is there no editorial oversight? Does it pretend to be a news site?
- VM: some sites just copy directly from major news sites. Having spoken to Grammarly about potential of identifying 'copying' sites or AI-generated sites, the detection technology appears not to exist.
- FS: when thinking about the impact of AI, we need to look beyond editorial rooms to the wider infrastructure around journalism. AI already being used for some boring tasks, such as identifying relevant files in archives.
- FS: you could have multiple versions of the same story for different audiences.
- VM: hopes labelling and transparency will become the norm. NewsGuard partners with Microsoft for Bing (see [here](#)), providing information about the quality of all sources.
- VM: AI is not going to be able to call for comment as most journalists can – *will it not?*
- FS: publicly funded news organisations to the resources and incentives to continue producing quality content.

- CJ: These organisations could become more valuable as a result of AI-generated journalism reducing the quality of other organisations' outputs. However, important to remember that AI is not the main problem for newsrooms: that's profit.
- VM: AI might help with boring/repetitive pieces of work.
- VM: Newsugard is testing chatbots' ability to produce misinformation in the style of certain sites (e.g., Fox News). Seems this makes it much easier to produce misinformation.
- VM: also warns about 'model collapse', the phenomenon whereby as models become increasingly powered by AI-generated content, their performance declines.
- GJ: publishers are so cash-strapped at the moment that they are giving up their archives to LLMs.

Generative AI, disinformation and extremism

Broderick McDonald (King's College London); Dina Hussein (Meta); Rayhan Asat (Atlantic Council); chaired by Esther Jaromi (QMUL)

- BM: scale and cost factors key, but don't overlook that generative AI can make content more immersive. Extremists are using gaming to recruit and set up communities – generative AI can change game code to make it more amendable to extremists' purposes. Extremists are also creating their own games.
- BM: extremists are using LLMs to set out their manifestoes in music, and these applications are also being used to write public (shooters') manifestoes. Video and audio of these can be made more compelling.
- BM: Generative AI also facilitates better translation – helping the far-right to become increasingly internationalists.
- DH: we shouldn't think of terrorist actors as being at the cutting-edge of this technology. Far-right are seen as most receptive to generative AI, as their content is (initially at least) based on reach and interactions, rather than overt ideology.
- DH: Meta using AI to upscale moderation. Trust and safety tools can identify uses of generative AI on its platforms.
- DH: cautions that mass surveillance can often be the consequence of intensive oversight of AI.
- BM: actors may claim incriminating clips are AI-generated. AI can also be used for detection.
- DH: AI can track and map networks very well.
- DH: red teaming of models essential.
- RA: authoritarian states deploy vast amounts of AI-generated disinformation against their own populations. Civil society can't compete against states alone, requires help from tech companies.

- DH: state actors are generally moving first with regard to new technologies. It takes time for the public to react and call for regulation. Meta conducts research and reaches out to the public, so that shared knowledge of these technologies increases. The risk is that we regulate in a way that certain actors/companies wish to be regulated.
- BM: *Tech Against Terrorism* has robust frameworks to identify terrorist sources – these sources must be classified as a terrorist organisation by multiple organisations (e.g., EU, UK, etc.), which prevents legitimate political enemies of the state being singled out.
- RA: community notes are actually bad, they turn into another vector for disinformation.
- DH: we need global best-practices approach. The ‘global’ is the hard part, but important to remember that individual companies don’t have all of the knowledge themselves.
- DH: response time for international bodies is too slow. For example, commissioning research and writing a peer-reviewed paper can take at least a year.
- DH: content generated by AI is still content, and existing content moderation laws still apply.
- RA: cautions against an expansive interpretation of national security to combat extremist AI-generated content.
- DH: when thinking about radicalism/extremism, we mustn’t forget offline solutions too. Not all about AI-generated content.

Social media company policies and overview

The purpose of this document is to outline key information about some of the main social media platforms which are currently used in election campaigns.

Included in this document is information on the following:

Contents

Overview	1
Snap	1
Facebook	2
Instagram	6
Threads.....	7
TikTok.....	8
Google / YouTube.....	10
X	12
LinkedIn	15

It does not include platforms not currently used for campaigning purposes in a widespread way, such as Reddit, Pinterest or Whatsapp, or smaller platforms like Mastodon or Telegram.

We will review which platforms are included on a regular basis to ensure this document captures the vast majority of UK online campaign activity.

Each section includes an outline of the functionality of the app, how it has typically been used by campaigners in election, its policies around political advertising, synthetic content and misinformation, and links to any call notes or engagement the policy team has had.

Overview

	Allows political advertising	Key definitions	Has an ad library	Has synthetic content policy	Misinformation/ fact checking policy
Snap	Yes. Additional restrictions on organic content posted on Discover and Spotlight.	<p>Election-related ads are defined as ads about candidates or parties for public office, ballot measures or referendums, political action committees, and ads that urge people to vote or register to vote.</p> <p>Issue or advocacy ads are defined as ads concerning issues or organizations that are the subject of debate on a local, national, or global level, or of public importance. They include ads about abortion, immigration, the environment, education, discrimination, and guns.</p>	Yes, although it is a basic excel file.	Same content moderation applies to synthetic content as human-created.	Spreading falsehoods that can cause harm is prohibited. Snap removes any content which is in breach of its guidelines.
Facebook	Yes. Advertisers need to be verified (have a UK-based passport and	Social issues, elections or politics ads applies to an ad that:	Yes. The most comprehensive example offered by platforms. However, there have been reported	Disclosure required on social issue, electoral, or political ads that contain a photorealistic image or video, or realistic	Removes misinformation where it is likely to directly contribute to the risk of imminent physical harm. Also removes

	Allows political advertising	Key definitions	Has an ad library	Has synthetic content policy	Misinformation/ fact checking policy
	residential address) and be admin of a Page.	<ul style="list-style-type: none"> • Is made by, on behalf of or about a candidate for public office, a political figure, a political party, a political action committee or advocates for the outcome of an election to public office • Is about any election, referendum or ballot initiative, including "get out the vote" or election information campaigns • Is about any social issue in any place where the ad is being run • Is regulated as political advertising 	discrepancies with the data included.	<p>sounding audio, that was digitally created or altered to:</p> <ul style="list-style-type: none"> • Depict a real person as saying or doing something they did not say or do; or • Depict a realistic-looking person that does not exist or a realistic-looking event that did not happen, or alter footage of a real event that happened; or • Depict a realistic event that allegedly occurred, but that is not a true image, video, or audio recording of the event. 	content that is likely to directly contribute to interference with the functioning of political processes and certain highly deceptive manipulated media. Partners with a network of verified factcheckers globally to monitor and fact check flagged content. Limits dissemination of spam/clickbait content.
Instagram	Yes. Same restrictions apply	Same as Facebook	Uses Meta library	Same as Facebook	Same as Facebook

	Allows political advertising	Key definitions	Has an ad library	Has synthetic content policy	Misinformation/ fact checking policy
	as with Facebook.				
Threads	Does not support ads	N/A	N/A	No	No
TikTok	No	No definition of 'political', but certain entities are prohibited from advertising on the platform, such as candidates, elected office-holders, and political parties.	N/A – doesn't allow political ads, but the platform allows you to search among the best performing ads.	Yes. Synthetic material depicting realistic scenes must be 'clearly disclosed'. Public figures can be depicted by synthetic media, but not private ones.	Misinformation about civic and electoral processes is not permitted. Has fact-checking processes, and a database for previously checked claims.
Google / YouTube	Yes	Election ads cover: <ul style="list-style-type: none"> • A political party, current elected officeholder, or candidate for the UK Parliament; or • A referendum question up for vote, a referendum campaign group, or a call to vote related to a national referendum or a regional referendum on sovereignty 	Yes	Yes. Political ads with synthetic content must prominently disclose when they inauthentically depict real or realistic-looking people or events. Manipulated media relating to politics that is designed to deceive users is also prohibited.	Prohibits misinformation that could significantly undermine trust in electoral and democratic processes. Collaborates with fact-checking organisations, attempts to boost authoritative content and limit the virality of content that may put users at risk.

	Allows political advertising	Key definitions	Has an ad library	Has synthetic content policy	Misinformation/ fact checking policy
X	Yes, since 2023. Advertisers need to be certified and ads are subject to additional eligibility, product, and targeting restrictions.	<p>Political content ads are defined as ads that reference a candidate, political party, elected or appointed government official, election, referendum, ballot measure, legislation, regulation, directive, or judicial outcome</p> <p>Political campaigning ads are defined as ads that advocate for or against, appeal directly for votes, or solicit financial support for</p> <ul style="list-style-type: none"> • a candidate or political party. • an election, legislation, referendum, or ballot measure. 	Used to have one but now is only US-based.	Bans users from sharing synthetic, manipulated, or out-of-context media that may deceive or confuse people and lead to harm (“misleading media”).	User-submitted Community Notes appear at the end of a post where the factual information provided has been called into question.
LinkedIn	No	Political content includes content such as political parties and candidates, election outcomes, and ballot initiatives. This does not include content on social	N/A	No	Bans misinformation and inauthentic content, as well as objectionable and harmful content.

	Allows political advertising	Key definitions	Has an ad library	Has synthetic content policy	Misinformation/ fact checking policy
		issues that may be discussed in a political context.			

Snap

Background on Snap

Snapchat was initially focused on private, person-to-person photo sharing, but you can now use it for a range of tasks, including sending videos, live video chatting, messaging, creating caricature-like Bitmoji avatars, and sharing a chronological “story” that’s broadcasted to all your followers.

Snapchat Has 750 Million Daily Active Users Worldwide, with 23.15 million users in the UK. The most popular age group on Snapchat is those ages 15-25, as they make up 48% of users.

There are several features on Snap. *Discover* displays branded content made by Vice, Cosmopolitan, Daily Mail, ESPN, CNN, BuzzFeed, and more. *Spotlight* is Snapchat’s answer to TikTok. It shows short-form video content, or "Spotlight Snaps", based on Snapchat's algorithm. Snap’s *My AI* is a chatbot which responds to prompts from users, similar to ChatGPT.

Political advertising

Snapchat does allow political advertising, although there are some restrictions under its [policy](#). Snap allows paid-for political advertising, but does not allow political content on its Discover or Spotlight pages (unless the content is newsworthy). Alongside this, there are additional moderation controls, both through proactive machine learning and human moderation. It does have an [ad library](#), but it is pretty basic (an excel spreadsheet).

Synthetic content

The same content moderation policies would apply whether the content was created by a human or AI. The main use of AI on Snap is its *MyAI* chatbot, which utilises OpenAI models. Snap confirmed at our last meeting it is currently looking at what additional guardrails were needed to ensure the AI did not generate political content.

Misinformation

Snap’s Community Guidelines prohibit spreading false information that can cause harm — including, for example, conspiracy theories, denying the existence of tragic events, unsubstantiated medical claims, or undermining the integrity of civic processes. Snap removes any content that goes against these guidelines.

Engagement



Facebook

Background

Facebook is a social media platform owned by Meta. Users can create a profile, post text, photos and multimedia which are shared with any other users who have agreed to be their friend or, with different privacy settings, publicly. Posts appear on user's timelines and newsfeeds, alongside posts from friends and any pages or groups they follow. Users can also create and join (private) Groups or (publicly accessible) Pages to share content with specific, interested audiences.

Facebook has made several changes to its platform in recent years. In 2018, Facebook announced that it would change News Feed to prioritize friends/family content and de-emphasize content from news companies. In 2017, Facebook launched *Facebook Stories*, as a competitor to Snapchat. Meta CEO Mark Zuckerberg announced that Meta would start selling blue "verified" badges on both Instagram and Facebook. Alongside this, *Facebook Messenger* allows instant messaging, voice and video calls.

As of December 2022, Facebook claimed 3 billion monthly active users.

How it is used by campaigners

Campaigners have made heavy use of Facebook in previous election campaigns. Political parties often set up Pages, with the larger parties setting up multiple Pages for each local party group. Pages or groups can also be set up by non-party campaigners, referendum campaigners, or individual candidates..

Campaigners can post both paid-for ads and organic content on Facebook. Facebook Ads can appear on all Facebook platforms including Instagram, Audience Network and/or Messenger. The available "Ad placements" vary depending on the platform.

Adverts can be targeted to specific types of users. Attribute-based targeting tools allow campaigners to manually select a target audience for a particular ad or ad campaign based on various characteristics, using data that the social media platform has previously collected and processed about individuals. Facebook lists five characteristics that can be selected for such targeting: (i) location, (ii) demographics, (iii) interests, (iv) behaviour, and (v) connections.

Personally Identifying Information (PII) tools allow campaigners to target their existing contacts via multiple methods. The most common way to execute this is by uploading a list of email addresses, phone numbers or mobile advertiser IDs campaigners already possess in order to then identify the associated social media accounts of the existing customer.

As well as paid content, campaigners can post organic content for free. This includes, publishing updates on their page, creating and promoting events such as rallies or hustings, encouraging people to vote..

Political advertising

The full policy as it relates to social issues, elections or politics can be found [here](#).

In order to post a paid-for ad, users need to be a Page admin where they wish to run the ad from. Facebook also requires a UK passport and a UK-based residential mailing address to verify advertisers. Advertisers must include a verified "Paid for by" disclaimer on these ads to show the entity or person responsible for running the ad.

Meta's [ad library](#) is the most comprehensive of all those offered by social media platforms – so for that reason it is the one we use the most in our monitoring work. People can search for all active ads running across products from Meta. For ads about social issues, elections or politics, Meta provides additional information, including spend, reach and funding entities. These ads are visible whether they're active or inactive and are stored in the Ad Library for 7 years.

However, Meta's library has come under scrutiny in recent years. For example, [this report from Ofcom](#) into monitoring political advertising found discrepancies in the data provided by libraries. For example, the list of advertisers from the Facebook Ad Library API did not exactly match the list of advertisers found in the Facebook Ad Library Report CSV files. These differences were not explained in any documentation and could potentially undermine the correct interpretation and use of the underlying data.

Meta's ad library includes ads that are defined as ads about social issues, elections or politics. This applies to any ad that:

- Is made by, on behalf of or about a candidate for public office, a political figure, a political party, a political action committee or advocates for the outcome of an election to public office
- Is about any election, referendum or ballot initiative, including "get out the vote" or election information campaigns
- Is about any [social issue](#) in any place where the ad is being run
- is regulated as political advertising

The inclusion of 'social issues' makes this definition is broader than what is used by other social media platforms and what is defined in PPERA or other parts of electoral law (such as the digital imprints regime). This means ads that aren't related to politics or elections often appear in the library – for example, any ad that mentioned COVID-19, regardless of the wider context of the ad.

Synthetic content

Since January 11th 2024, a new policy [introduced by Meta](#) will mean advertisers will have to disclose whenever a social issue, electoral, or political ad contains a photorealistic image or video, or realistic sounding audio, that was digitally created or altered to:

- Depict a real person as saying or doing something they did not say or do; or
- Depict a realistic-looking person that does not exist or a realistic-looking event that did not happen, or alter footage of a real event that happened; or
- Depict a realistic event that allegedly occurred, but that is not a true image, video, or audio recording of the event.

This policy only applies to paid-for advertising (not organic). On the organic side, [manipulated media is listed](#) as a category of misinformation which is removed by Meta.

According to its policy, Meta removes videos under this policy if specific criteria are met:

- the video has been edited or synthesised, beyond adjustments for clarity or quality, in ways that are not apparent to an average person and would likely mislead an average person to believe a subject of the video said words that they did not say; and
- the video is the product of artificial intelligence or machine learning, including deep learning techniques (e.g. a technical deepfake), that merges, combines, replaces and/or superimposes content onto a video, creating a video that appears authentic

Misinformation

Meta's full policy on misinformation can be found [here](#).

Rather than blanket-banning 'misinformation', Meta articulates different categories of misinformation with different approaches to each.

According to their policy: "We remove misinformation where it is likely to directly contribute to the risk of imminent physical harm. We also remove content that is likely to directly contribute to interference with the functioning of political processes and certain highly deceptive manipulated media. In determining what constitutes misinformation in these categories, we partner with independent experts who possess knowledge and expertise to assess the truth of the content and whether it is likely to directly contribute to the risk of imminent harm.

For all other misinformation, we focus on reducing its prevalence or creating an environment that fosters a productive dialogue. Finally, we prohibit content and behaviour in other areas that often overlap with the spread of misinformation. For example, our Community Standards prohibit fake accounts, fraud and coordinated inauthentic behaviour."

This policy [replaces](#) the earlier iteration which prevented candidates from running ads arguing during the US 2020 election was stolen.

Co-ordinated Inauthentic Behaviour / Foreign influence

Meta has [this policy](#) on **Inauthentic Behaviour**. This bans people from misrepresenting themselves on Facebook, using fake accounts, or artificially boosting the popularity of content. In particular, users cannot conceal a Page's purpose by misleading users about the ownership or control of that Page, use multiple Facebook accounts or share accounts between multiple people

Engagement



Instagram

Background

Instagram is a photo and video sharing app, owned by Meta. Posts can be shared publicly or with pre-approved followers. Users can browse other users' content by tags and locations, view trending content, like photos, and follow other users to add their content to a personal feed. Since launch, it has added messaging features, the ability to include multiple images or videos in a single post, and a *Stories* feature—which is similar to Snapchat, where photos and videos appear in carousel format for 24 hours before disappearing.

In August 2020, Instagram began a pivot to video, introducing a new feature called *Reels* to compete with TikTok. Instagram also added suggested posts in August 2020, where posts from accounts not followed but related to the user's interests are displayed on feeds. There are currently over 2.4 billion monthly active Instagram users, 68% of which are female while 32% are male. 90% of the 150 million users are under the age of 35.

How it is used by campaigners

Campaigners can make use of Instagram's free features by posting organic content on their page or on Stories. This includes pictures, videos or Reels. They can also post paid-for advertising through Meta and these ads will appear in the Meta political ad library. In recent years, Instagram has seen substantial political material, especially from non-party campaigners and pressure groups – especially around specific issues, for example #BlackLivesMatter and #MeToo.

Political advertising

Ads can be [posted](#) and managed either directly on Instagram or through Facebook. The same political advertising policies apply on Instagram as on Facebook.

In addition, political actors, such as candidates and elected officials are not permitted to make use of branded content. This is defined as creator or publisher's content that features or is influenced by a business partner for money. This means they're not allowed to make money directly from posting on Instagram, unlike other influencers.

Ad library

Ads that fall under the above definition will appear in the Meta ad library.

Misinformation

Instagram uses the same fact checking programs as Facebook. Since 2017, Instagram has employed the ability to reduce the prominence of accounts ("shadowbanning") it believes may be generating non-genuine engagement and spam. Instagram's strict content moderation policies around nudity have been criticised by some users.

Threads

Background

Threads is a social media app owned by Meta. Its interface is very similar to that offered by the microblogging app, X. Users can create posts consisting of up to 500 characters (compared to Twitter's 280 characters), as well as videos and photos. However, Threads lacks other common social media features, such as hashtags, trending stories, and direct messaging

The app is closely linked with Instagram, and requires users to have an Instagram account and use Threads under the same handle.

Launching in July 2023 in 100 countries, it initially gained over 100 million users in its first five days. However, its early success was not sustained and the user base of the app plummeted more than 80% to 8 million daily active users by the end of July.

How it is used by campaigners

It is very rarely used by campaigners.

Political advertising

There are currently no plans to monetise Threads – so there are currently no ads on the app – including political ads.

Content moderation/fact-checking

Threads has the same community guidelines as Instagram. It has [received](#) some criticism in the press over the current lack of proper controls on content, specifically around election participation disinformation.

TikTok

Background

TikTok is a video-sharing app owned by the Chinese company ByteDance, primarily used on mobile phones. Users can upload short videos, and have access to video editing and customisation tools. In addition to being a key vector for the proliferation of internet 'memes', TikTok is increasingly playing a greater role in online shopping, hosting third-party sellers who embed their products into videos.

Concerns have been raised about connections between ByteDance and the Chinese state, with the Director of the FBI [cautioning](#) that users' personal data and devices could be compromised by the app. TikTok is not permitted in China, although ByteDance operates another almost identical app there, called Douyin.

Recent figures suggest that TikTok has at least 1.9bn users (and 15.9m in the UK), making it one of the most popular social media platforms in the world. Its growth has been rapid: before ChatGPT and Threads, it was the quickest app to reach 100m daily users from its launch date. Its users tend to skew younger, with relatively similar numbers of men and women using the app (although this varies between countries).

TikTok has a number of features designed to encourage users to engage with existing trends, including allowing them to tap on and copy audio used in other clips, as well as duet with other users' videos. Users can toggle between viewing videos on their 'For You' page, which algorithmically suggest videos based on their interaction with other videos, or their Friends or Following pages.

How it is used by campaigners

While TikTok's terms and conditions prohibit political advertising on the platform, and ministers have been [banned](#) from using the app on their government phones, a number of MPs use TikTok to interact with the public and with constituents. These include: [Grant Shapps](#), [Matt Hancock](#), and [Zarah Sultana](#). Content posted by politicians include interventions in Commons debates, constituency work, and details about their personal lives.

In *The Times*, Katy Balls has [predicted](#) that both parties will increasingly turn to TikTok, following the importance of the platform in other, international elections. Parties are likely to increasingly employ influencers on TikTok to boost their messages: despite TikTok's claim that political advertising and receiving payments to create political content are against its rules, there are multiple [examples](#) of this on the platform.

Political advertising

Political advertising is not allowed in any form on TikTok, covering both TikToks being promoted, or paying for individuals to make political content. Users can create organic political content with a political message.

TikTok does not set out definitions for what it considers 'political', but its [policies](#) are outlined below (notably this makes no mention of non-party campaigners):

- Candidates or nominees for public office, political parties, and elected or appointed government officials are prohibited from advertising.
- The spouses of candidates, elected, or appointed government officials with official duties or offices are prohibited.
- Royal Family members with official government capacities are also prohibited.
- Government entities may be eligible to advertise if working with a TikTok Sales Representative.

Although TikTok has an ad library, as it theoretically does not allow political ads, it can be difficult to find them in the library. The library itself allows you to filter by popularity (by likes); by recency (last 7, 30, and 180 days); the industry posting the ad; and the purpose of the ad (e.g., traffic, app installations, etc.).

Misinformation

TikTok's [Community Guidelines](#) state that 'inaccurate, misleading, or false content that may cause harm' is not allowed on its platform. The definition of 'harm' does not extend to commercial or reputational harm, but it does include undermining democratic elections. Misinformation about civic and electoral processes (how to vote, voter registration, election outcomes, counting of ballots, and the eligibility of candidates) is not permitted on TikTok, regardless of intent.

TikTok partners with [independent fact-checkers](#) (although there doesn't appear to be a UK-specific one), and has a database of previously decided claims that are used in any decision-making. An internal investigation and moderation team work with the fact-checkers to identify any election-related misinformation.

Misinformation about civic or electoral process can lead to several [actions](#), including: removal of content; redirecting search results and hashtags to TikTok's Community Guidelines; reducing the virality of content; removal of the account; or even removal of the device from the platform.

Synthetic content

TikTok's [community guidelines](#) require that synthetic content depicting realistic scenes is 'clearly disclosed'. Synthetic material purporting to show private individuals, as opposed to individuals with a 'significant public role', is not allowed. Users may not use synthetic material to provide endorsements.

Engagement



Google / YouTube

Background

Google is the world's most popular search engine, accounting for over 80% of all worldwide users. Google Search (the most popular website in the world) uses algorithms to rank web pages based on their relevance to the user's query. In addition to Google Search and other Google-branded products, Google Inc. owns a number of other major companies, including DeepMind and YouTube (the latter also being an important source of political advertising).

How it is used by campaigners

Campaigners can advertise on Google and its products in three main ways: search ads, whereby certain results are promoted to users based on their searches; YouTube ads, which can be tailored to users' watching history; and display ads, which appear on websites.

Political advertising

Google does allow political advertising, but attaches certain conditions. Advertisers may only run election adverts in the UK if they have been [verified](#), a process which requires proof of an organisation's name and address (if they choose to register as an organisation – the process is slightly different for individuals, who can also run political adverts) and attestation.

'UK Election ads', as Google calls them, cover any advert [featuring](#):

- A political party, current elected officeholder, or candidate for the UK Parliament; or
- A referendum question up for vote, a referendum campaign group, or a call to vote related to a national referendum or a regional referendum on sovereignty

Google does have a [public ad library](#), but as *Who Targets Me* has [noted](#) (presumably about Google search ads, rather than YouTube ads), there are substantial ways in which it could be improved. For example, it could capture political adverts outside of election periods, and it could expand the definition (see above) of political adverts to cover local election-related adverts or adverts by non-party campaigners.

Adverts posted on YouTube and hosted in Google's ad library contain information about the amount spent on promoting the ad (a range is given, rather than a precise figure), how many times it was shown (again, a range), and targeting criteria (limited to age, gender and location for political ads).

Misinformation

Google [claims](#) that it does not allow content that ‘makes claims that are demonstrably false and could significantly undermine participation or trust in an electoral or democratic process’. This includes information about voting procedures, the eligibility requirements for candidates, or the results of elections. The company [notes](#) that such content may be blocked, removed, or restricted. However, in 2022, [ProPublica](#) found Google continuing to place ads in violation of this policy.

Google also [bans](#) directing users to content about politics in a country outside of the publisher’s, if the publisher misrepresents or conceals details about their identity or country of origin.

Google also [has](#) a team of experts who ensure that users are exposed to high-quality information, boosting authoritative sources, limiting the spread of some content, and removing content that puts people at risk. However, information about the work of this team is quite limited. The company has a [Fact Check Explorer](#), containing fact checks about certain people and topics.

Synthetic Content

In late 2023, Google [announced](#) that political ads in the UK (and other countries where verification is required to post election ads) must ‘prominently disclose when their ads contain synthetic content that inauthentically depicts real or realistic-looking people or events’. Such disclosures should be clear and prominent. Smaller edits, or inconsequential AI-generated elements of the advert, need not be disclosed.

Google also has a longer-standing [manipulated media](#) policy, which prohibits content that ‘deceives users through manipulated media related to politics’, among other things. In the Commission’s discussions with Kelly Thomas from Google, it appears that the distinction between the two policies lies in intent.

Engagement



X

Background

X (formally known as Twitter) is a micro-blogging site and social media platform. In October 2022, it was acquired by Elon Musk for \$44bn. On X, users can share and post text messages, images, and videos known historically as "tweets". X also includes direct messaging, video and audio calling, bookmarks, lists and communities, and Spaces, a social audio feature. Posts (or tweets) are publicly visible by default, but senders can restrict message delivery to only their followers. Users may subscribe to other users' posts—this is known as "following". Users can mute users they do not wish to interact with, block accounts from viewing their posts, and remove accounts from their followers list.

X Premium (formally Twitter Blue) provides additional services, including adding the blue tick on accounts and the prioritisation of content on other users' For You pages.

Since the acquisition, the platform has been criticized for increasing the spread of [disinformation](#), [antisemitism](#), and [transphobia](#). Analysis [conducted](#) by research firm Sensor Tower found that global active daily users of the platform via mobile apps had steadily declined during the year after Musk acquired the company, down 16% by September 2023. Several [advertisers](#), including Apple, Lionsgate and Disney have pulled their advertising on the platform. By December 2023, loss of advertising revenue had become so extreme that public media speculation suggested that Twitter would become bankrupt.

In 2020, it was estimated that approximately 48 million accounts (15% of all accounts) were not genuine people.

How it is used by campaigners

X has become widely used in campaigning, with world leaders, politicians, government departments, as well as non-party campaigners, candidates, parties and pressure groups all making use of the app. However, due to the previous ban on political advertising, most recent activity by campaigners is organic. This includes posting rallying support by posting tweets, sharing content such as infographics, commentary or news articles, or engaging with other accounts' tweets through commenting, retweeting or liking.

Chief executive Linda Yaccarino has told industry figures she was aiming for the platform to bring in \$100mn annually in political ad revenues in 2024. However, [some analysts](#) have said the targets are grossly overambitious and that X lends itself more to soliciting donations, collecting email addresses and efforts to boost voter turnout, rather than shifting opinions, as users of the platform tend to be entrenched in their political beliefs.

Political advertising

In August 2023, X [announced](#) it was lifting its ban on political ads on its site. The Political Content policy means advertisers must obtain pre-approval to run political ads by first getting certified.

Advertisements are subject to additional eligibility, product, and targeting restrictions:

- Political campaigning ads targeted to a country can only be bought by citizens or entities based out of that country.
- Advertisers must comply with the country specific election laws (including campaign and silent periods).
- All advertisers must comply with [X's Sensitive Categories Targeting Policy](#). Only the following criteria may be used to target political campaigning ads: age, location, gender, interests and keywords, custom audiences, follower look-alikes.

X requires companies to participate in the [verified organisations program](#) in order to purchase advertising on the platform, although companies that spend at least \$1,000 on advertising per-month automatically receive membership in the program at no additional cost.

There is currently no ad library for the UK – the only library appearing currently on the site is for US- and EU-based ads.

Synthetic content

X's [synthetic and manipulated media policy](#) bans users from sharing synthetic, manipulated, or out-of-context media that may deceive or confuse people and lead to harm (“misleading media”).

In order for content with **misleading media** (including images, videos, audios, gifs, and URLs hosting relevant content) to be labeled or removed under this policy, it must:

- Include media that is significantly and deceptively altered, manipulated, or fabricated, or
- Include media that is shared in a deceptive manner or with false context, and
- Include media likely to result in widespread confusion on public issues, impact public safety, or cause serious harm

Memes, satire, animations, commentary, reactions and “counterspeech” are not excluded from this policy.

However, X has a high threshold for proving if something is a deepfake – it needs to be 100% proven as such to be covered. For this reason, the [audio of Kier Starmer](#)

allegedly abusing a staffer was not taken down, although Community Notes were added to it explaining that some believed the video to be fake.

Misinformation

Community Notes [aim](#) to address mis/disinformation on X. These are user-submitted notes that appear at the end of a post where the factual information provided has been called into question. Anyone can sign up to be a contributor. If multiple contributors with opposing views (judged on previous engagement with notes) rate a piece of content as useful, then the note is judged by the algorithm as accurate.

Engagement



LinkedIn

Background

LinkedIn is a social media platform focused on business and employment. Since December 2016, it has been a wholly owned subsidiary of Microsoft.

The platform is primarily used for professional networking and career development, and allows jobseekers to post their CVs and employers to post jobs. LinkedIn allows members (both workers and employers) to create profiles and connect with each other in an online social network. LinkedIn can also be used to organise offline events, join groups, write articles, publish job postings, post photos and videos

There are, since February 2022, approximately 310 million LinkedIn users.

How it is used by campaigners

LinkedIn has in recent years seen an increase in number of politicians creating accounts and using it for self promotion and campaigning purposes. For example, Rishi Sunak, Kier Starmer and Ed Davey all have pages, as do all the major political parties and non-party campaigner groups. Although the platform does not allow political advertising, campaigners can create organic content.

In 2022, LinkedIn [added](#) the ability for users to toggle to a 'no politics' feature. This uses machine learning to identify and remove political content, such as posts from political figures and politically-gearred comments.

Political advertising

Political ads are prohibited under [LinkedIn's advertising policies](#).

This includes ads advocating for or against a particular candidate, party, or ballot proposition or otherwise intended to influence an election outcome; ads fundraising for or by political candidates, parties, political action committees or similar organizations, or ballot propositions; and ads exploiting a sensitive political issue even if the advertiser has no explicit political agenda.

Synthetic content

LinkedIn does not have a synthetic content policy.

Misinformation

Misinformation is [banned](#) on LinkedIn, as is inauthentic behaviour. It has several schemes members can take advantage of to help them identify misinformation, including the [News Literacy Project](#), [The Trust Project](#) and [Verified](#).

AI Safety Summit – Notes

Date: 1-2 November 2023

Bletchley Park Mansion

- Day 1 of the Summit focused on roundtable discussions about the key risks concerning AI and what governments and the international scientific community can do to mitigate these risks. A brief summary of the discussions can be found [here](#): the positive and negative effects of AI on elections and democracy was raised in ‘Roundtable 4’, but there’s little substantive in the note.
- Day 2 of the Summit involved the Prime Minister Rishi Sunak convening a small group of governments, companies and experts, while the Secretary of State for Science, Innovation and Technology Michelle Donelan discussed future steps with the UK’s international partners.
- The Summit produced four main outputs:
 1. The [Bletchley Declaration](#) (see below for more).
 2. The [production](#) of a ‘State of the Science’ report reviewing the capabilities and risks of frontier AI. The report’s production will be chaired by Yoshua Bengio, who has been one of the foremost advocates of controls on the development of AI. Each country present at the Summit will [nominate](#) experts to help with the writing of the report.
 3. An [agreement](#) between the US and UK to work on separate AI Safety Institutes, which will exchange analyses. More information on these institutes can be found in this [article](#).
 4. Companies who possess many of the leading AI applications have also [agreed](#) to allow the governments of 10 countries (including the US, UK, Japan, France, and Germany) to review their products before they are publicly released. This will be done in part by the UK’s AI Safety Institute (see ‘Commentary’ section).

[Bletchley Declaration](#)

- The Declaration (signed by all attendee countries) encourages parties to continue identifying and sharing risks around frontier AI, and to develop domestic risk-based policies in response to these risks. ‘This includes, alongside increased transparency by private actors developing frontier AI capabilities, appropriate evaluation metrics, tools for safety testing, and developing relevant public sector capability and scientific research.’
- The Declaration explicitly mentions ‘unforeseen risks stemming from the capability to manipulate content or generate deceptive content’ and disinformation, although there is no mention of elections or democracy, despite the fact that elections were on the agenda for the first day of the Summit.

- As was reported in the run-up to the Summit, the main focus of the Declaration is on ‘frontier AI’, although there is some reference to the current use of AI and the sensible principles which should guide its usage.
- It seems that the AI Safety Summit will become an annual event with a rotating head (*as Chloe Smith mentioned at the Oxford Generative AI Summit*), with South Korea hosting a smaller virtual summit on AI in six months’ time.
- The Declaration is full of warm words about future collaboration, but there appear to be no new mechanisms to achieve this.

Other

- Ahead of the Summit, a number of AI companies (including OpenAI and DeepMind) [published](#) their safety policies, in response to a request from the Secretary of State Michelle Donelan.
- In his interview with Elon Musk, the Prime Minister acknowledged the risks around AI-related disinformation at elections, and trailed [plans](#) to use an AI chatbot to help people pay taxes and access their pensions.

Commentary

- During the Summit, Peter Kyle MP (Shadow Secretary of State for Science, Innovation and Technology) [reiterated](#) Labour’s call for stronger regulation of frontier AI, which would include independent oversight of certain models’ safety and a requirement for companies to report whether they are training models over a certain capability. He expressed concern that if left unchecked, the most powerful AI models could spread misinformation, undermine elections, and help terrorists build weapons.
- While obtaining both Chinese and American support for the Bletchley Declaration was seen as a diplomatic coup, the [Guardian](#) reported that there was some disappointment that attendees had not agreed to expand the capabilities of the UK government’s AI taskforce. American Vice-President Kamala Harris used her trip to the UK to [announce](#) the establishment of a United States AI Safety Institute, as competition grows over which country should be at the centre of setting global AI regulations.
- Trade unions, human rights organisations, and academics signed an open [letter](#) to the Prime Minister, expressing disappointment that the Summit did not focus on the impacts of AI on ‘the communities and workers most affected by AI’. Vice-President Harris made mention of a wider range of AI-related concerns in her [speech](#) at the US Embassy, including risks to workers and to democracy.
- Beyond the Declaration, there has been a recent flurry of high-level pronouncements on the future use of AI. These include: the aforementioned White House executive [order](#); a G7 [code of conduct](#) for companies; and the EU’s AI Act. The UN also has an [AI advisory board](#) which is expected to report before the end of the year.
- The Summit received broadly positive reviews from a number of [news outlets](#).