Can artificial intelligence be used to undermine elections?

The UK and the US are both heading towards national elections. In the next couple of years almost every major democracy will also hold the most important national elections in their systems. Previous elections in the UK and the US and a referendum in the UK saw spikes in political disinformation online, as well as online exercises in profiling and targeting for the purpose of political influencing. Similar activities have been identified around elections worldwide.

The online influencing that happens around these coming elections will not simply be repeats of what we have seen before. Since the last US and UK general elections, artificial intelligence has continued to develop fast. New applications including large language models have been developed, and some have become widely available. It is very likely that some of these new tools will be used in attempts to influence the functioning and the outcomes of those elections. Some of those uses may fall within the range of safe and legitimate political activity. Others may be harmful and even dangerous to democratic processes.

Collectively, we should improve our understanding of what could go wrong and how we should be protected from new harms, before the elections take place.

How could AI applications threaten elections?

AI tools might be used to affect elections in several ways, including these.

- Creating and spreading misinformation and disinformation
- Undue manipulation of public opinion
- Targeting voters with personalised content without informed consent
- Hacking election systems
- Disrupting civil society around elections

In practice, these different modes could well be used in combination. As well as increasing disruption, that could lead to cumulative loss of confidence in the election process and its results.

None of these types of threat are new. All have featured in attempts to disrupt elections in the past. However, AI applications may have potential to multiply the effectiveness, volume and frequency across all these categories of threat. Where generative AI technologies (that themselves create and distribute new information and expressions of information) are involved in particular, potentially harmful applications are becoming available to a very wide range of individual actors. A Russian propaganda technique that has been called the “firehose of falsehood” involves broadcasting messages at speed and through many channels. AI applications may make similar effects more widely achievable.

Increasing use of some AI tools may also add challenges to traditional modes of oversight, regulatory action and explainability in relation to online techniques for influencing, for instance making it difficult to detect and justify why a particular content item was shown to a particular individual at one time and in one online context.

Creating and spreading misinformation and disinformation: AI applications can be used to create news articles, social media posts, photos, soundbites and videos designed to mislead voters and influence their voting choices or decisions to vote at all. This category covers a wide range of types of online content, from mistakes, misattributions and misinterpretations, across a spectrum to orchestrated campaigns using AI applications to generate realistic-looking deepfake video.

There is plenty of content to draw on to make new disinformation appear credible. There is also plenty of past disinformation to inform developers on what tropes tend to get the most traction.

This category overlaps with targeting, as content could be tailored to the beliefs and biases of individual voters or groups. These types of content may be intended to achieve specific changes in voting intention, but at the same time have a broader objective in poisoning public discourse and wearing down confidence in it.

As well as making it easier to lie, widespread public expectation of deepfakes may create a "liar's dividend: a means for political figures to dismiss real evidence against them, described by Robert Chesney and Daniella Keats Citron.

> As the public becomes more aware of the idea that video and audio can be convincingly faked, some will try to escape accountability for their actions by denouncing authentic video and audio as deep fakes. Put simply: a skeptical public will be primed to doubt the authenticity of real audio and video evidence.

Undue manipulation of public opinion: AI applications could be used to manipulate public opinion and behaviour online in other ways, by creating or reinforcing echo chambers in online fora, where people are only exposed to information that confirms their existing beliefs. This can make it more difficult for voters to make informed choices.

Research on filter bubbles has suggested that risks from them have sometimes been overestimated. People can be exposed to a range of opinions and materials online, that leave them better informed and more able to identify disinformation. Newer applications may increase or accelerate the capability to cause groups to align, form and act collectively.

As we will explore below, this is an area where it can be hard to define exactly what is acceptable. Political campaigning is intended to influence opinion, and should be allowed to do so. However, techniques that mislead people about the opinions of others and encourage polarising separation into opposing online groups based on disinformation may fall beyond the limits of what is acceptable.

Targeting voters with personalised political content without informed consent: Social media can be used to target voters with political ads that are specifically tailored to their interests and demographics. Records of previous activity could be used to influence voters' choices, especially if the ads are designed to exploit their fears or prejudices. AI applications could increase the effectiveness of the direction of content to individuals based on data on what would influence them, and even at what time of day they might be most open to influence. Bespoke chatbots could frame interactions aimed at voter characteristics, changing and honing messages, at relatively low costs to those aiming to influence.

Hacking election systems: AI applications could be used to hack into online systems for voter registration and for voting (where those are in use). Attacks could extract personal data for subsequent targeted actions (for instance of people likely to use alternatives to voting in person). Attacks on voting systems could nullify, alter or misreport votes. AI-enabled cyberattacks could potentially bring down whole election systems, delaying or making impossible accurate announcement of outcomes.

Attacks on systems present a serious threat to the integrity of elections. There is a secondary threat that attacks on parts of the system can undermine confidence in the whole and give credibility to allegations about the integrity of the process.

Disrupting civil society around elections: AI-enabled cyber-attacks could potentially be directed against a very wide range of targets that support the civil society in which elections happen. Cyber-attacks might be directed to create traffic jams, power outages, or other obstacles that make it difficult for people to vote. This would suppress voter turnout but might also be targeted to certain places and populations, skewing the impact on election outcomes in favour or against some parties and candidates.

In fact, all of these categories of interference have potential uses not just in changing how people vote, but in changing who votes, in what is known as voter suppression. Particular social and local groups could be targeted with techniques and messages that seek to misrepresent candidates they might be expected to vote for, confuse them with conflicting messages, discount the value of voting at all, and falsely undermine the probity and security of the election system.

Countering AI interference in elections

Comprehensively addressing these threats may require effective collaboration between governments, law enforcement organisations, media, civil society organisations, and the tech industry. There are already ways to address disinformation and manipulation online, some of which may benefit from updates to address the specific challenges posed by use of AI applications. Recognition of risks by governments is growing, as expressed in a 2019 Declaration by the Committee of Ministers of the Council of Europe.

> Fine grained, sub-conscious and personalised levels of algorithmic persuasion may have significant effects on the cognitive autonomy of individuals and their right to form opinions and take independent decisions. These effects remain underexplored but cannot be underestimated.

Governments and regulators should explore ways to promote transparency specifically in the use of AI in elections.

Social media companies are major channels for the spread of disinformation online, and Governments already require them to take action to remove particularly dangerous categories of illegal content. Governments could increase requirements on these companies to take additional steps to prevent the spread of disinformation, removing fake accounts and posts, labelling misleading content generated by AI applications, and critically examining targeting techniques.

In August 2023 the US Federal Election Commission announced that it had launched a rule-making process on AI deepfakes for political purposes. Election law already prohibits the fraudulent misrepresentation of candidates and campaigns. This change, if taken through, would "make it clear that the related statutory prohibition applies to deliberately deceptive Artificial Intelligence (AI) campaign advertisements" as well.

It may be appropriate to require labelling of auto-generated political content online and of the use of AI otherwise in political communications and campaigns. The White House recently announced seven "Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI". One of these was a commitment "to developing robust technical mechanisms to ensure that users know when content is AI generated, such as a watermarking system." Any such systems would require careful monitoring of actual effects in society as AI becomes more ubiquitous. These are unlikely to resolve all risks from autogenerated disinformation, but should make a difference.

Google has already begun to deliver on this commitment, launching a tool that "embeds a digital watermark directly into the pixels of an image, making it imperceptible to the human eye, but detectable for identification." A valuable step in itself, this should also encourage parallel and competitive action from others.

What may be uncertain right now is how effective and resilient platforms' processes will prove in the face of the greatly increased quantity and variety of online disinformation which may be enabled by more recent technology developments. Part of the challenge is achieving clear, consistent and impactful application of terms and conditions in relation to novel content. Part is about resources: keeping up with the quantity of misleading material, which tends to peak before elections. Part is technical: improving tools for detecting, labelling and removing content and identifying bad actors as they improve their tools for creating convincing content.

Detecting disinformation is a literal race to act fast enough on dangerous content before it has effects. From the longer perspective it is also an unremitting arms race. A recent European Commission report stated that "in many cases the mitigation measures introduced by online platforms failed to account for the Kremlin's malign intent and full scope of information warfare tactics employed on online platforms."

Social media companies have developed their capabilities in this regard, but not as much as many civil society organisations would like them to do. For government to act legitimately to induce the platforms to do more, we may need a more urgent and thorough conversation about what we role social media should play in democratic societies.

Governments can invest in election oversight and security (including cybersecurity), and in the skills and capability of independent regulators for elections and data protection and use. They can invest in training of election officials on recognising and addressing emerging threats.

Political parties could commit to clear shared commitments on how they will use new technologies in campaigning. Public debate could better establish a collective understanding of what is acceptable practice and what is not.

Governments can increase support for “independent, evidence-based and interdisciplinary research and advice for decision-makers regarding the capacity of algorithmic tools to enhance or interfere with the cognitive sovereignty of individuals”, as the Council of Europe put it. Governments can collaborate more internationally to improve shared understanding of threats, and they can share good practice in protecting elections. Some of this is currently left to non-governmental organisations with limited and intermittent funding.

The online public, the users of platform services, can do more to protect themselves and democratic processes, if they are sufficiently aware of the risks of AI-enabled disinformation. A 2022 Ofcom report on Adults’ Media Use and Attitudes suggested that some people do not recognise when online content is misleading, and others overestimate their own ability to do so.

- There was often a gap between people’s confidence in being able to recognise advertising, identify a scam message or judge the veracity of online content, and their ability to do this when shown examples.
- A third of internet users were unaware of the potential for inaccurate or biased information online; 6% of internet users believed that all the information they find online is truthful and 30% of internet users don’t know – or don’t think about – whether the information they find is truthful or not.

A general, thorough education and up-skilling across society is needed to ensure that power to recognise and deal with threats is distributed as any concentrated power is a potential threat to democracy. Governments and online service providers could do more to educate the public about how new threats could manifest and be recognised as such. Helping voters make informed choices could include teaching them how to spot misinformation, or how to protect their personal data from being used by political campaigns.

The rapid evolution of online tools can help as well as threaten. Online applications nudge users to check facts or direct them to alternative information sources, as Richard Mackenzie-Gray Scott has recently explored.

> Digital nudges have the potential to reduce the spread of misinformation on these platforms and may well be the least rights-intrusive means of doing so when compared to alternatives (with the possible exception of user reporting).

However, as he also explains, these rely on the same kind of techniques used for spreading misinformation and manipulating opinion, and can easily run up against principles and law protecting freedom of thought and opinion.

Why is this difficult to address?

There are constraints on governments' ability and willingness to act in this area. Some of these are also important to democracy, so there are real tensions. Other constraints have less principled causes, but are still relevant.

Governments in liberal democracies are rightly concerned with protecting free speech online, and can be uncomfortable about being seen to curtail it or to impose standards of what is acceptable political activity. Authoritarian governments often constrain or switch off internet traffic around elections. Many democratic governments want to avoid even the appearance of doing anything similar.

Even with the best intentions, regulating to suppress some kinds of political expression can involve making distinctions that are difficult to define or to apply consistently. Some material can be easy to condemn, for instance wholly fictional misleading deepfake material. Some misleading online content may be much harder to judge as going beyond what should be protected as free expression. Legislative processes are designed to create regulation that works in practice, that can as far as possible be easily understood by the public and by companies who need to follow and sometimes implement it. It is not straightforward to write rules that deliver acceptable balance between freedom of speech and public protection, when the range of what is possible changes quickly.

There are generic obstacles to regulating to prevent misuse of general purpose technologies which also have many legal and beneficial uses. Governments can make specific misuses illegal, but effective preventive measures are always more difficult where it is not possible or reasonable to limit any access to the technology. Communications technologies can fall into that group, and the internet is perhaps the supreme example. Open source large language models may be another class.

In terms of new tools including generative AI, proprietary and open source present different challenges to potential regulation. Open source tools are more available for misuse, but it may be easier to anticipate and address the variety of potential misuse. Proprietary tools offer less transparency, but may in other respects be easier to focus regulation and governmental influence on.

Governments are led by politicians who have a strong interest in the freedom to make political arguments and to influence voters, using all legitimate tools available. Many politicians like campaigning, and relish the opportunity to show that they are better at it than the competition. For some, that includes showing that they are better at using new tools and channels. That can be part of a pitch to appear more modern and technocratically competent.

Even relatively small changes by governments to election law and to election institutions can become the focus of suspicion, and that suspicion may often be justified. Elected politicians may not be the best arbiters of what is appropriate political campaigning. Even where they set rules, it may be better in terms of transparency and public confidence to keep them at a distance from applying them.

Major internet platforms present special challenges for governments, in relation to maintaining a public information sphere. Liberal democracies are constitutionally opposed to regulating collective public access to information. Unrestricted marketplaces for information and opinion are sometimes seen as the most important of free markets. Governments have been reluctant to take strong positions on online platforms' intermediary role in transmitting information, where that does not involve acts or material that clearly break existing laws.

At the same time, online social media platforms create peculiar challenges to maintenance of a shared public reality. Everyone does not see the same material. Typically, social media platforms are designed to affect user behaviour online, to increase time spent online, and to give people what they want to see or are engaged by seeing. The business model relies on the capacity to target users with adverts that they are individually more likely to be influenced by, using accumulated information about individuals to customise their experiences, and to increase the capability of the influence over time. The platforms aim to do this without making the influencing or the personalisation too overt, so parts of the system are effectively hidden from the targets.

Platforms are reluctant to give up any part of this capability, or disclose more about it than they are obliged to, because achieving personalisation at scale in targeted advertising is a key source of competitive advantage for them. Even though much internet activity is funded on the capability of online advertising to change behaviour, there are (perhaps surprisingly, given the scale of the businesses and markets) many questions about online influencing: how it works, how well it works, and when and why it works particularly well, and what it (so far) cannot do. People and their representatives cannot seek redress from harms they cannot perceive or evaluate. Governments and regulators have not had access to much of the relevant data held by the global internet platforms.

Meanwhile, governments can be reluctant to admit that they do not fully understand what is happening in important markets. They may also be reluctant to admit that they are uncertain about the potential seriousness of emerging threats, for fear of being accused of sleeping on the job.

Perhaps we should not expect governments to have very clear ideas about how to regulate increasing capability for online influencing for political purposes, when they have not formed firm positions, principles or regulatory tools in relation to the increasing power of online influencing for commercial purposes.

The focus of existing legislation may leave gaps, in particular around collective interests. Most data protection law establishes individual rights to prevent harms to individual people. These have already arguably proved to have limitations either for protecting single citizens from the strategies of very large companies, for addressing the growth of data-enabled power in internet markets, or for addressing collective social impacts.

The House of Commons Science, Innovation and Technology Committee has recently published its interim report on the governance of artificial intelligence. In an evidence submission to the Committee, Dr Steve Rolf gave a relevant example.

> The draft AIA [draft EU Artificial Intelligence Act], for instance, includes outright bans on decision-making algorithms in cases where they pose a threat to ‘safety, livelihoods and rights of people’. While this places important emphasis on individual rights (somewhat absent in China's approach) it does little to address societal-level harms. An illustration of such harms might be impacts on democratic processes — for example, algorithmic recommendations on social media platforms that discourage wavering voters from turning out, thus tipping the balance in an election.

In relation to political influencing and protecting elections in particular, there is a need for better information and guidance based on global experiences. Given the number and variety of online actions to influence elections worldwide, much remains unknown about what kind of campaigns achieve influence, and about what works in countering inappropriate activity. Even around the better known examples including Russian interference in the 2016 US election and the Brexit referendum, and the actions of Cambridge Analytica, there is still uncertainty about how much outcomes of election processes were affected.

This is changing. Recently, four papers were published with results from a major research project involving researchers from Meta and several institutions examining the influence of Facebook and Instagram's algorithms on political news exposure in the US 2020 election. High level conclusions were that algorithms were extremely influential in users’ on-platform experiences, there was significant ideological segregation in political news exposure on Facebook, but that algorithms that determine what users saw did not sway political attitudes.

This kind of research, which is likely to be further examined and challenged, should ultimately make for a better shared understanding of how the internet intersects with collective political life.

A particular knowledge gap right now in governments and the public sphere is around just what is coming next: how much more effective tools for influencing online behaviour have become since 2016. The expectation, extrapolating along a path-dependency for the major platforms, is that improvements will have been made possible by more data, compute power and innovations, including innovations driven by AI. But what those improvements have been, and what they mean for the power to influence online, are not clear from outside major technology companies. Increases in capability may not have been linear. Creators of tools almost certainly do not know every purpose they can applied to. This is all moving very fast, as was pointed out in a recent report from the Brennan Center, “How AI Puts Elections at Risk”.

> The launch of ChatGPT just weeks after the 2022 midterm election on November 30, 2022, has precipitated a new era in which many people regularly converse with AI systems and read content produced by AI.

Even that is beginning to seem some time ago, in the scale of the generative AI era.

The major technology companies and their moderation policies are also subjects of political observation and critique, while they are active lobbyists themselves. They are frequently accused of censoring content from different sources disproportionately and in line with perceived political biases. Recent research has suggested different large language models can be placed on a map of political predispositions. In AI ethics, there has been a divide between those pointing to current problems resulting from use of AI applications in ways that affect the public now, and those looking to the further future. Predilection to one or other view of where to focus also appears increasingly to fall across older political divides.

It is challenging in this context to establish the political objectivity which we would want from organisations acting to protect democracy. This should improve the case for augmented independent expert oversight and regulation of elections. There should be more clarity on what uses of personal data are legitimate for political purposes, in particular in relation to emerging technologies, and around elections.

Risks to elections can be different in form and potential impact in different states. States that have relatively weaker election security and oversight, or lack civil society safeguards including established independent media, may be subject to different and potentially more impactful orchestrated online campaigns.

They may also be used as experimental spaces. It is already evident that techniques for influencing an election in one country have been trialled, honed and subsequently reapplied in other countries. Collectively, all democracies should benefit from shared research, combined with informed assessments of developing capabilities offered by AI that help to imagine what could come next.

Personal data and UK elections: an update

In this context, the electorate could reasonably demand clarity in law relating to how personal data is used politically online, and stability, confidence and independence in the organisations responsible for oversight of elections. However, a lot has happened recently in this space.

Legislative changes in the UK have given political parties more rather than less scope in their use of personal data. The 2018 UK Data Protection Act included new exemptions to some restrictions on use of personal data. One of the amendments specifically permits political organisations to make use without consent of “personal data revealing political opinions”, where that “is necessary for the purposes of the person’s or organisation’s political activities” including “campaigning, fund-raising, political surveys and case-work”.

One obvious problem here is that this is a new provision, untested around a UK general election, so we do not know how political parties will make use of it. In a time when technologies for influencing online are evolving quickly, that is not reassuring.

The exemption does not give political parties an entirely free hand. The Information Commissioner's guidance of lawful basis elaborates.

> Many of the lawful bases for processing depend on the processing being "necessary". This does not mean that processing has to be absolutely essential. However, it must be more than just useful, and more than just standard practice. It must be a targeted and proportionate way of achieving a specific purpose. The lawful basis will not apply if you can reasonably achieve the purpose by some other less intrusive means, or by processing less data.

This should put pressure on political parties to assess and justify their processes. However, that justification might only be sought after an election has happened. Given the potential for using novel methods relying on newer technologies, there is an argument for requiring more explanation and justification around what political parties are doing with personal data, to regulators if not in public releases.

There are potentially good arguments for political parties having access to certain classes of information about voters. If they know more about citizens, they should be able to better develop policies to understand and address their needs. However, political parties should have those rights only through collective consent, following well-informed national debate informed by full and shared understanding of what might be done with the information with available and emerging technologies.

The government's current Data Protection and Digital Information Bill includes new provisions supporting the use of unsolicited direct email for political objectives, and enables the government to make future regulations to enable direct marketing for democratic engagement.

Other changes made and proposed relate to election process and oversight. The Data Protection and Digital Information Bill also proposes to reduce the independence of the Information Commissioner's Office (the UK's independent regulator protecting information rights) remove the Surveillance Camera and Biometrics Commissioner, and reduce the requirements for organisations using the personal data of UK citizens to have independent Data Protection Officers and undertake Privacy Impact Assessments. The 2022 Elections Act placed the Electoral Commission, the body which oversees elections and regulates political finance in the UK, under the supervision of a government minister. The Commission had previously been independent of government and accountable directly to parliament. Civil society organisations have criticised the changes, both those made and those proposed.

Also in recent news, the Electoral Commission announced on 8th August that it had been the target of a complex cyber-attack in which "the perpetrators had access to the Commission's servers which held our email, our control systems, and copies of the electoral registers." The incursion was identified in October 2022 after suspicious activity was detected, and it became clear that hostile actors had first accessed the systems in August 2021. The registers to which access was gained included the name and address of anyone in Great Britain who was registered to vote between

2014 and 2022, of overseas voters, and of voters registered in Northern Ireland in 2018.

The Commission acknowledges risks to the public from subsequent use of the data in combination with data from other sources.

> According to the risk assessment used by the Information Commissioner's Office to assess the harm of data breaches, the personal data held on the electoral registers – typically name and address – does not in itself present a high risk to individuals. It is possible however that this data could be combined with other data in the public domain, such as that which individuals choose to share themselves, to infer patterns of behaviour or to identify and profile individuals.

Among its other responsibilities, the Commission works "to promote public confidence in the democratic process and ensure its integrity." Undermining that confidence may have been a key objective of the attack, aside from subsequent use of the data. Many commentators have asked why the Commission waited months to announce that the attack had taken place.

Many of these developments on their own may have small or ephemeral effects on individuals or in the short term, but together they may leave UK citizens concerned about what may be done with their personal data to influence their voting decisions. At the moment, there is a worrying of well-informed public debate about how untested changes to relevant laws and organisations could affect delivery of elections, in an environment of serious and growing technological challenges.

Political parties could seek to reassure the public by making disclosures and binding commitments about how they propose to use citizens' data in forthcoming elections, and in particular how they intend to develop and use AI applications with that data. In line with the Information Commissioner's guidance, they could explain in advance of elections why the use is "a targeted and proportionate way of achieving a specific purpose". This could be a matter for public discussion before the event, and not only for the work of a small number of poorly-funded investigative journalists afterwards.

In May, the Government published further proposals to amend the Elections Act, which included confirmation that a new online registration service is in development. That should be subject to extensive public scrutiny, given the growth in potential threats.

Where do we go from here?

The UK will only realise full benefits of AI if we can collectively develop more confidence in the governance and regulation of new technologies. That confidence also must be justified and periodically renewed. Developers of a new wave of applications propose that machines will be able to read our minds soon. We should not have to live in fear of the future, so we need to update protections for freedom of thought and opinion, and freedom to freely choose elected representatives, so these established democratic necessities work under new technological conditions.

Democratic elections are part of what enable society to adapt to the future. We should be able to have confidence that they work fairly, even if we sometimes individually dislike the results. We should demand better accountability, oversight, and transparency around the potential effects of AI systems on elections. We should be reassured that elections are fully protected from threats that will continue to evolve.