Artificial intelligence, democracy and elections

SUMMARY

Artificial intelligence (AI) has become a powerful tool thanks to technological advances, access to large amounts of data, machine learning and increased computing power. The release of ChatGPT at the end of 2022 was a new breakthrough in AI. It demonstrated the vast range of possibilities involved in adapting general-purpose AI to a wide array of tasks and in getting generative AI to generate synthetic content based on prompts entered by the user. In a just a few years' time, a very large share of online content may be generated synthetically.

AI is an opportunity to improve the democratic process in our societies. For example, it can help citizens to gain a better understanding of politics and engage more easily in democratic debate. Likewise, politicians can get closer to citizens and eventually represent them more effectively. Such an alignment between citizens and politicians could change the face of electoral campaigns and considerably improve the policymaking process, making it more accurate and efficient.

Although concerns over the use of AI in politics have been present since the late 2010s, those related to democracies and the election process in particular have grown with the recent evolution of AI. This emerging technology poses multiple risks to democracies, as it is also a powerful tool for disinformation and misinformation, both of which can trigger tensions resulting in electoral-related conflict and even violence. AI can, for example, generate false information, or spread a bias or opinions that do not represent the public sentiment. Altogether, despite its benefits AI has the potential to affect the democratic process in a negative way.

Despite the above risks, AI can prove useful to democracies if proper safeguards are applied. For example, specific tools can be employed to detect the use of AI-generated content and techniques such as watermarking can be used to clearly indicate that content has been generated by AI. The EU is currently adapting its legal framework to address the dangers that come with AI and to promote the use of trustworthy, transparent and accountable AI systems.

IN THIS BRIEFING

EPRS | European Parliamentary Research Service

Author: Michael Adam with Clotilde Hocquard
Members' Research Service

Introduction

While concerns about the use of artificial intelligence (AI) in politics and its impact on the democratic process are not new, they have been provoked further by the rise of general-purpose AI and generative AI, which represents a technological breakthrough. Ninety per cent of online content may be generated synthetically by 2026. Although there is no scientific consensus on how to define AI, the European Commission's Joint Research Centre has come up with an operational definition whereby AI systems are seen as software systems designed by humans that decide the best action to take to achieve a certain goal, by acquiring data and processing the information derived from this data. These AI systems are commonly used to provide personalised recommendations to people based on their previous searches or other online behaviour. In November 2022, AI research company OpenAI made available to the public its chatbot ChatGPT, a generative AI system designed to generate text after a human user enters a prompt. Generative AI systems are usually trained on large language models that use deep learning algorithms to learn and adapt without following explicit instructions and to draw inferences from patterns in data.[1] Generative AI systems can then replicate and predict similar patterns to generate content. Since the release of ChatGPT, multiple rival generative AI solutions capable of generating text, images and sound have entered the market. Generative AI is expected to become even more powerful in the coming months, and video generators based on text prompts will look even more authentic in the near future. ChatGPT also belongs to the category of general purpose AI, which encompasses AI systems trained on large language models to be adapted to a wide range of tasks across different domains.

The functioning of those new AI systems is relatively opaque and information about how the data is collected or trained is often unavailable. On top of privacy and intellectual property concerns, AI has a potential for bias, manipulation and spreading of disinformation, which risks weakening societies. However, if proper safeguards and rules on transparency are implemented, AI can be an extremely useful tool to improve the democratic process, particularly during election times.

AI as a catalyst of democracy

AI tools to improve political engagement

As a UNESCO study underlined, AI has the potential to improve democratic values, institutions and processes, including elections in various ways.

AI can serve to educate citizens in the principles of democratic life, whether by gaining knowledge about a policy issue or getting familiar with a politician's stance. For instance, political recommender systems could form the basis of a chatbot responding to citizens' questions on candidates' electoral programmes.[2] Moreover, specially designed AI tools could update citizens on how policies in which they have an interest are evolving[3] and empower them to better express their opinions when addressing governments and politicians. Civic debate could improve thanks to the capability of AI to manage massive political conversations in chat rooms. AI could automatically summarise participants' opinions, moderate the debate by identifying tensions and nudging them away from attacks and insults, and even act as a consensus builder.

On the politicians' side, AI can be helpful in summarising citizens' comments made during public consultations or received by email. Feedback could be classified according to various criteria, helping politicians gain a better understanding of citizens' views, especially if combined with human expertise.[4] Afterwards, politicians could use AI to draft personalised replies to the citizens. A study showed that such use of AI does not undermine citizens' trust as long as careful disclosure of the use of AI is made and human oversight is guaranteed.[5]

New forms of election campaigns

The way political campaigns are run may also see major changes with the rise of AI. Politicians will soon have the capacity to use AI to respond instantly to campaign developments. The increased use of sentiment analysis on social media will also allow them to better understand the topics citizens are sensitive to and interested in. AI could then be used to generate personalised emails or text messages from chatbots to specific audiences. The other side of the coin is that such an understanding allows to very precisely target one's audience, meaning that politicians could eventually target specific groups of voters, among them swing voters.[6]

A transformative effect on the policymaking process

AI could also play an important role in the policymaking process and generate more value in each of its five stages: identification, formulation, adoption, implementation and evaluation. AI's ability to summarise complex problems and to process vast amounts of data can help policymakers to identify societal issues. AI can also deepen policymakers' knowledge of and provide democratic assemblies with expertise on broad topics, thus sparing them the difficulties associated with restricted availability or access. On the other hand, AI could eventually be used for new forms of lobbying, by summarising a proposal for new legislation, assessing whether it is of relevance to a company and even drafting a letter to its author arguing for changes to the proposal.[7] AI could also help select key legislators and target them through direct communication and public relations campaigns. Researchers have shown that AI could be used to enact micro-legislation, which involves making small modifications to proposed legislation that would have the biggest impact on a narrow area of interest.[8] While AI could democratise lobbying activities for individuals and non-profits by reducing their cost, there is a risk that AI capabilities may influence public policy towards outcomes that do not reflect societal aspirations in a fair way.

Disinformation and misinformation fuelled by AI

AI equips malicious entities with a wide variety of techniques to influence public opinion. First, AI can help to observe the information environment and understand the emerging social fissures. The network analysis capabilities of AI can also be used to better target an audience and establish the profile of voters, in what is known as political micro-targeting. AI can dramatically increase the speed at which content is made, while also offering access to a wealth of resources. Consequently, this could give rise to entire fake-news websites posing as news outlets. New AI tools also make it possible to generate images from text or to clone a person's voice. Deepfake videos are getting easier to produce and are becoming more and more convincing, to the point that text-to-video is described as the upcoming breakthrough in generative AI.

Deepfakes have a huge potential for misinformation (false or inaccurate information), or even disinformation (information having as its intention to mislead), notably through the making of memes and humorous video content, both of which often go viral online. Politicians are the main potential target of deepfakes, especially when they do not have the resources to protect their online presence.[9] Overall, deepfakes severely risk undermining trust in the information environment. They also may make it easier for some politicians to dodge responsibility for their real words, on the pretext of having fallen victim to AI-generated content.

The breakthrough in generative AI raises concerns regarding influence campaigns, as it will now require less human and financial resources to conduct large-scale disinformation campaigns. The OECD stressed that the combination of AI language models and disinformation can lead to deception on a large scale and damage public trust in democratic institutions.[10] AI systems themselves can produce misinformation. A study found that Google's Bard AI tool generates persuasive misinformation content on 78 out of 100 tested narratives. The new version of ChatGPT, ChatGPT-4, is even more susceptible to generating misinformation and more convincing in its ability to do so than its predecessor, ChatGPT- 3.5.

Although AI providers are working on making their models more reliable, such as by running red-team exercises, AI can easily be used to draft and spread harmful narratives online, directly tailored to country-specific contexts. Disseminating disinformation could become even more impactful with the use of interactive chatbots, which can customise interactions based on voter characteristics, but also adapt manipulation tactics in real time and apply them to a multitude of citizens. Such AI models could be conceived as anthropomorphised tools and generate content that simulates human emotions to manipulate the user. Generally, AI presents an important manipulative potential, as users may not be able to distinguish between human and AI-generated content. Researchers demonstrated AI's power of persuasion by showing that across different topics, AI-generated messages were at least as persuasive as human-generated messages, and that users are even more likely to trust tweets generated by AI than content written by humans.

Negative impact of AI on democratic engagement

Politicians usually perceive mails and correspondence from their constituents as an expression of public opinion on which they can act. The rise of AI makes it possible to conduct astroturf campaigns, where a small group, posing as a genuine grassroots movement, presents a skewed view of public opinion. For instance, AI could be used to generate false correspondence with the aim of influencing legislators. An experiment showed that legislators found AI-generated text they received on six policy areas almost as credible as human-written messages. AI could also be used to create the illusion of political agreement, by posting millions of automatically generated content entries on a topic online. It could further be used to compose comments on regulatory processes, or to write letters to the editors of local newspapers, which are then published.

Tools to counteract harmful impacts of AI

The globalpolicy.ai coalition, gathering organisations from around the globe, has been discussing the possibility of launching a Global Challenge on Digital Trust to incentivise stakeholders around the world to develop solutions to mitigate the risks of disinformation. Currently, several methods can help deter AI-generated content from featuring in disinformation campaigns online.

Automated detection tools

First, guardrails can be implemented by AI systems providers against certain types of speech, but they can easily be circumvented, for instance by feeding ChatGPT-3 with fictitious scenarios in prompts, by asking it to take on the role of a character, or by making slight modifications to the spelling of keywords. Moreover, automated detection tools, such as GPTZero, OpenAI's classifier and DetectGPT, can flag AI-generated content but are not entirely reliable, as they are quickly rendered obsolete due to AI's increasing ability to generate more fluent language.[11] OpenAI's own tool correctly identified only 26 % of AI-written texts in the evaluations the company launched, although it has proven to be more accurate with longer texts. Detection tools can also easily be circumvented by replacing a few words with synonyms, or by paraphrasing AI-generated text. To improve the efficiency of detection tools, some experts advocate for open-source AI models, which would make it easier for the forensics community to build effective detection tools for tackling deepfakes, among others.

Watermarking and innovative tools to identify AI-generated content

Another technique called watermarking involves the addition of a visual label to indicate that a content is AI-generated, or to mark the metadata. For instance, the image generator DALL-E incorporates a colourful stripe in its images. Google also announced that it would attach a written disclosure underneath AI-generated results on Google Images. Such a watermark is however sometimes hard to spot and can easily be removed. When it comes to AI-generated text, the provider can watermark the output by embedding patterns into the model's word choice that do not substantially change the output meaning. Consequently, they remain invisible to humans but are algorithmically detectable. The limitations of this method are that it may reduce the quality of AI-generated text due to its vocabulary constraint and that the watermarks can be removed by paraphrasing the AI-generated text.

It is worth noting that the industry has been working on creating an interoperable open standard called C2PA to label AI-generated content and disclose its provenance or ownership history. Adobe for example is now using this standard as part of its content credentials tool, enabling creators to add extra information about themselves and their creative process directly to their content at export or download.

Other solutions that are being brought forward to counter AI-generated disinformation include the development of an instant fact-checker for social media users, the introduction of human detection of AI-generated content, or the identification of networks of fake accounts on social media platforms. Although distinguishing text generated by AI as opposed to humans is a difficult task, it remains necessary especially for fact-checking breaking news. The EU is currently funding a number of projects to build trustworthy AI solutions, one such project being vera.ai, which includes a fact-checker-in-the-loop approach (where the AI model is re-trained continuously with the help of the users' feedback) and AI models that constantly check updated and verified sources.

EU legal framework

AI can both benefit and endanger democracy, as was already highlighted above. EU law already contains some rules to address the risks linked to the use of AI tools. For instance, the General Data Protection Regulation (EU) 2016/679 (GDPR) and the EU Data Protection Regulation (EU) 2018/1725 give users the right to object to profiling but also restrict profiling based on the use of sensitive personal data. In addition, the EU is currently adapting its legal framework to better address the risks of AI, including for democracy.

Code of Practice on Disinformation

The European Commission launched a series of non-legislative initiatives to tackle disinformation, starting with a communication presented in 2018. There followed the adoption of a Code of Practice on Disinformation, where industry – including major online platforms – voluntarily agreed on self-regulatory standards to fight disinformation. In 2021, the Commission issued guidance to strengthen the Code of Practice on Disinformation, and in 2022, a new Code of Practice on Disinformation was adopted. The industry agreed to put in place stronger transparency measures on political advertising, by providing more efficient labelling, committing to reveal the sponsor and the advertising spend and display period, and to put in place searchable advertising libraries for political advertising. The code also aims to empower users by making available tools to recognise, understand and flag disinformation and to access authoritative sources, and by running media literacy initiatives.

Digital Services Act

The Digital Services Act (DSA) entered into force in November 2022. Under the DSA, very large online platforms have to take a risk-based approach through independent audits of their risk management systems to prevent abuse – such as disinformation – of their systems. They then have to take action to mitigate the risks, including by moderating the content displayed on their platforms.

The Code of Practice on Disinformation is recognised under the DSA. Consequently, for very large platforms, complying with the code could be considered an appropriate risk-mitigating measure.

The DSA further addresses concerns regarding the micro-targeting of citizens. It bans both targeted advertising to minors based on profiling, and targeted advertising based on profiling using special categories of personal data such as political opinions. It also imposes transparency requirements for advertising on online platforms.

AI can be a powerful tool to improve content moderation and detect fake news on social media. However, reducing the visibility of content also entails risks of blocking legitimate forms of expression, limiting the circulation of legitimate content, restricting democratic debate and limiting pluralism during electoral periods. The risk of bias that can come with AI models can reinforce these concerns. To address the problem of platforms' opacity and understand their impact on societies, the DSA enables vetted public interest researchers to access the data of very large online platforms and thus enhance their accountability to the public.

Political advertising

The European Parliament and the Council of the EU are currently holding trilogue negotiations on the proposal for a regulation on political advertising. The proposal seeks to harmonise the rules on the transparency of political advertising by making record-keeping and labelling of political advertisements obligatory. Labels would have to be accompanied by a transparency notice mentioning the sponsor's identity and contact details, the period of publication and the amounts spent. The proposal would also ban targeting and amplification techniques that involve the processing of sensitive personal data such as political opinions.

AI Act

Moreover, the proposed AI act envisages a risk-based approach and imposes specific regulatory requirements on high-risk AI systems. In its position on the proposed AI act, the Parliament added to the list of high-risk AI systems those used to influence voters in political campaigns. It also introduced a layered approach to regulating general-purpose AI. Providers of foundation models (defined in the Parliament's position as AI system models that are trained on broad data at scale, are designed for generality of output, and can be adapted to a wide range of distinctive tasks), would have to ensure robust protection of fundamental rights, democracy, the rule of law, health, safety and the environment. Furthermore, generative foundation AI models, aimed at generating text images, audio or video, would have to disclose the fact that the content was generated by AI and not by humans. Moreover, they would have to be trained and designed to prevent generation of illegal content. Institutional negotiations are currently ongoing to finalise the proposed AI act.

MAIN REFERENCES

Ghost in the machine, addressing the consumer harms of generative AI, Norwegian Consumer Council, June 2023.

Madiega T., General-purpose artificial intelligence, At a Glance, EPRS, March 2023.

Elections in digital times: a guide for electoral practitioners, UNESCO, 2022.

Sedova K., McNeill C., Johnson A., Joshi A., Wulkan I., AI and the Future of Disinformation Campaigns, Part 2: A Threat Model, Center for Security and Emerging Technology, 2021.

Monteleone S., Artificial intelligence, data protection and elections, At a Glance, EPRS, 2019.

ENDNOTES

[1] EU-U.S. Terminology and Taxonomy for Artificial Intelligence, first edition, Trade and Technology Council, May 2023, p. 9.

[2] Schneier B., Farrell H. and Sanders N., How Artificial Intelligence Can Aid Democracy, April 2023.

[3] Artificial Intelligence and Electoral Integrity, Concept Paper, European Conferences of Electoral Management and Bodies, 2022, Part IV 'How can AI enable a better informed voter choice and a higher turnout?'.

[4] Dooling B. and Febrizio M., Robotic rulemaking, April 2023.

[5] Bender S., 'Algorithmic Elections', *Michigan Law Review*, Vol. 121, No 3, 2022, p. 34, December 2022.

[6] West D., How AI will transform the 2024 elections, May 2023.

[7] ibid.

[8] Sanders N. and Schneier B., How AI could write our laws, March 2023.

[9] What a Pixel can tell : text-to-image generation and its disinformation potential, Disinfo Radar Project, September 2022, p. 30.

[10] AI language models – Technological, socio-economic and policy considerations, OECD Digital Economy Papers, April 2023, p. 10.

[11] Heikkilä M., Why detecting AI-generated text is so difficult (and what to do about it), February 2023.

DISCLAIMER AND COPYRIGHT

This document is prepared for, and addressed to, the Members and staff of the European Parliament as background material to assist them in their parliamentary work. The content of the document is the sole responsibility of its author(s) and any opinions expressed herein should not be taken to represent an official position of the Parliament.

Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged and the European Parliament is given prior notice and sent a copy.

© European Union, 2023.

Photo credits: © Emiliia / Adobe Stock.

eprs@ep.europa.eu (contact)

www.eprs.ep.parl.union.eu (intranet)

www.europarl.europa.eu/thinktank (internet)

http://epthinktank.eu (blog)