

ETHICS OF AI AND DEMOCRACY: UNESCO RECOMMENDATION'S INSIGHTS

This article focuses on the impact of Artificial Intelligence (AI) on democracy. It provides an overview of the various ways in which the AI technologies affect the democratic values and processes and the social and political behaviour of citizens. It notes that the use of AI, and its potential for abuse by some governments, as well as by big private corporations, poses a real threat to the institutions, processes, and norms of rights-based democracies. As governments worldwide mull over their AI strategies and policies that would protect citizens against AI-powered dangers, UNESCO announced a remarkable consensus agreement among 193 member states creating the first-ever global standard on the ethics of AI that could serve as a blueprint for national AI legislation and a global AI ethics benchmark.

Gabriela Ramos*



* Gabriela Ramos is the Assistant Director-General for the Social and Human Sciences of UNESCO. Previously, she was OECD Chief of Staff and Sherpa to the G20/G7. The author would like to thank Roy Saurabh (Consultant, UNESCO's Section for Bioethics and Ethics of Science and Technology Section) for his input and assistance on the publication of this draft.

Democratic values should be safeguarded and strengthened in the digital world. Artificial Intelligence technologies have already shown immense transformative power in almost all verticals impacting humanity. Democratic values and processes are no exception. AI technologies are not neutral and encode the values of the creators and underlying development and implementation ecosystem. They can be deployed to strengthen accountability within public institutions, elected leaders and can produce many benefits for citizen action, participation, and pluralism, making democracy more inclusive and responsive. However, they can also be used to strengthen autocratic capabilities and be leveraged for potentially malicious, manipulative purposes. Indeed, the rapid integration of AI technologies into public-facing applications and digital platforms provides unique opportunities for targeted, individualized, and often invisible influences on individuals and social groups, which different political stakeholders may be tempted to use to their own benefit.¹ There is a growing consensus that artificial intelligence (AI) will be a determining factor for the future of humanity, and AI is already influencing the critical functions of democracy (e.g., interference in electoral processes, personalized political targeting, shaping voters' behaviours and spreading misinformation to manipulate public opinion).² The article notes that concentration of information and related insights in the hands of a few big private actors, beyond democratic oversight, is a cause for concern.

To counter this, UNESCO's Recommendation on Ethics of AI calls for setting up solid national and international regulatory frameworks to ensure democratic governance of AI and prevent its misuse. The recommendation focuses on establishing transparent, accountable, and understandable AI ecosystems, that protect human rights. The ethical use of AI also underscores the role of algorithms in social media platforms and their possible implications for democracy. The regulatory framework should be based on values and principles and calls for more co-operation between private companies and international organizations. The developers and deployers of AI-technologies have so far favored a self-regulation stance in this field. However, self-regulation has proven, thus far, not sufficient in addressing these challenges and in protecting human rights, democracy, and rule of law. UNESCO, through the Recommendation, finds itself in a strategic position to provide the necessary guidance and support, in close co-operation and co-ordination with relevant institutions and organizations, towards creating a global regulatory framework for Artificial Intelligence that deals with the whole life cycle of AI systems, is addressed

¹ Anthony Nadler, Matthew Crain and Joan Donovan, *Weaponizing the Digital Influence Machine: The Political Perils of Online Ad Tech, Data & Society Report*, (17 October 2018): p. 47. <https://datasociety.net/library/weaponizing-the-digital-influence-machine/>

² "AI will be a Determining Factor for the Future of Humanity, Committee hearing is told," *Parliamentary Assembly, Council of Europe*, 4 October 2019. <https://pace.coe.int/en/news/7654>

to all stakeholders, and includes mechanisms to ensure the implementation of this instrument.

This article focuses on how AI influences and impacts the functioning of democracy, and how multiple stakeholders can engage in and contribute to the dialogue on AI. Above all, it makes a case for creating a common ground where institutions and private companies can establish close co-operation to build a shared democratic AI governance framework.

“Moreover, the broad use by governments of AI technologies to control citizens such as automated filtering of information amounting to censorship, and mass surveillance using citizen devices and public infrastructure coupled with vast integrated databases, may lead to the erosion of political freedoms and the emergence of authoritarian regimes that are powered by AI.”

Impact of AI on Democracy

Democracy, by design, should provide safeguards against the concentration of power in the hands of a few and can function properly only if based on credible, robust institutions and underlying processes that enjoy confidence of engaged, committed, and informed citizens and its ability to provide for dynamic balance of interests of relevant constituents. The crisis of modern democracies touches almost all elements of democratic order, including erosion of, and loss of confidence in institutions, mis- and disinformation plaguing public communications, fault lines in societal cohesion and increasing polarisation of communities. Modern technologies, including AI systems could both help resolve and aggravate this crisis.

AI technologies can be used to strengthen government accountability and can produce many benefits for democratic action, participation, and pluralism, making democracy more direct and responsive. However, it can also be used to strengthen repressive capabilities and for manipulation purposes. Indeed, the rapid integration of AI technologies into outreach mainstream communication tools and social media platforms provide unique opportunities for targeted, personalized and often unnoticed influence on individuals and social groups, which different political actors may be tempted to use to their own benefit.

The experience of the last few years helps to identify some key areas where the use of AI technologies can threaten to undermine and destabilize democracy. This article explores these key areas in detail that include access to information (misinformation, “echo chambers” and erosion of critical thinking), targeted manipulation of citizens, interference in electoral processes, erosion of civil rights, shifts of financial and political power in the data economy. Moreover, the broad use by governments of AI technologies to control citizens such as automated filtering of information amounting to censorship, and mass surveillance using citizen devices and public infrastructure coupled with vast integrated databases, may lead to the erosion of political freedoms and the emergence of authoritarian regimes that are powered by AI. Some observers call this “digital authoritarianism”—a new social order competing with democracy.³

Access to Information

Democracy requires well-informed citizens and implies that people with diverse perspectives come together to discuss to find common solutions through dialogue. By determining which information is shown and consumed, AI technologies used in digital platforms and other channels can contribute to advancing misinformation and hate speech, create “echo chambers” that lead individuals into a state of intellectual isolation where there is no place for dialogue, thus eroding critical thinking and disrupting democracy. Also, by prioritizing the news and information which users like, with commercial or other purposes, algorithms tend to reinforce their opinions, tastes and habits, and limit access to diverging views, thus reducing users’ free choice.

When it comes to the role of algorithms in advancing misinformation and hateful speech, content moderation algorithmic rules influence visibility, i.e., to what extent algorithms can identify and suppress posts that break community standards and cross the line when it comes to spreading bad/false information. Several AI-based platforms exercise automated censorship of content published on social media and deny or take offline information and views that the owners of platforms dislike, thus restricting freedom of expression. These embedded algorithms determine what individual users see online, including user-generated or organic posts and paid advertisements. Some of the most visible examples include Facebook’s News Feed, Twitter’s Timeline, and YouTube’s recommendation engine. These powerful, opaque algorithms also determine which users should be shown a given advertisement. The advertiser usually sets the targeting parameters (such as demographics and presumed interests), but the platform’s algorithmic systems pick the specific individuals who will see the advertisement and determine its placement

³ European Parliament. Directorate General for Parliamentary Research Services., *The Ethics of Artificial Intelligence: Issues and Initiatives* (LU: Publications Office, 2020). <https://data.europa.eu/doi/10.2861/6644>

within the platform.⁴

So, despite many technology platforms arguing that they are pursuing a hands-off policy regarding content by simply allowing users to say what they would like and not interfering with their free speech rights, they are silently putting their hands on the scale to determine which posts will be viewed and read by millions, i.e., which posts will go viral. The lack of responsibility allocated to the platforms by law, creates this vague space where content published is not submitted to the same rigor as traditional editorial boards. Thus, their algorithms are very much morphing what users see and what users react to. At present, considering that the overriding incentive that private digital platforms follow is revenue and profit, it can be assumed that even when content spreads misinformation, the algorithm will bump up its visibility, if it increases user engagement on the site.⁵

“AI technologies are fuelling a potent threat to democracy in form of an unprecedented and mostly un-checked concentration of data, information, and power in the hands of a small group of major digital companies which develop and own the algorithms, as well as the centralization of the Internet itself.”

Precise Targeting

Political partisan discourses, polarizing campaigns, and distortion of ‘facts’ to fit popular opinions are not new instruments in the political toolbox but AI technologies have tremendously amplified their scale and outreach. Thanks to AI, digital platforms play an increasingly important role in the political process to influence people and to further or block partisan interests. Some trends, already widespread globally, include large scale co-ordinated misinformation, including through deep fakes, micro-targeting of voters, polarisation of public debate, undermining confidence in democratic institutions, political parties, and politicians, as well as public trust in the reliability of information, control of information flow and public opinion.

During elections, AI can be effectively used to engage the voters on an individual level throughout the entire election process. Chatbots and discussion forums

⁴“It’s Not Just the Content, It’s the Business Model: Democracy’s Online Speech Challenge,” *New America*, accessed 16 February 2022. <http://newamerica.org/oti/reports/its-not-just-content-its-business-model/>

⁵ Emmanouil Papadogiannakis et al., “Who Funds Misinformation? A Systematic Analysis of the Ad-Related Profit Routines of Fake News Sites,” *ArXiv:2202.05079 [Cs]*, (10 February 2022). <http://arxiv.org/abs/2202.05079>

on social media platforms encouraging people to leave feedback at the end are all various ways in which the public mood can be gauged. Moreover, AI can help collect all this data in real time and enable party campaigners to alter their campaigns, accordingly, depending on what the public feels about them. In addition, AI can be used to manipulate individual voters. By analyzing the unique psychographic and behavioural user profiles of voters, AI is being used to persuade people to vote for a specific candidate or even create a bias against that candidate's opponent, and to strengthen the certainty about their choice. Psychographic profiling of citizens and targeting married with big data, deployed in digital political campaigns based on deception and coercion, influence activities from propaganda, persuasion, policy making. The same is clearly evidenced in a politically important case study - the actions of now-defunct political data analytics and behaviour change company, Cambridge Analytica, in the UK's 2016 referendum campaign on leaving the European Union.⁶

While micro-targeting for political campaigns may simply be seen as commercial advertising by digital platforms, it may threaten democracy, public debate, and voters' choices substantially when the related practices rely on the collection and manipulation of users' data (big data analytics) to anticipate and influence their political opinions and election results (computational propaganda).

Erosion of Civil Rights

Data availability and rapid progress in AI systems are witnessing an increased use of predictive analytics, not only by companies, banks, and recruiters, but also by government institutions and authorities. If the related shortcomings and risks are not addressed adequately, the AI-based amplification of bias and prejudice, as well as statistical flaws and errors, could lead to an entrenchment of historical inequity. This could undermine protection from discrimination and guarantees of equal treatment, which are enshrined in the constitutions of modern democratic societies.⁷

AI systems' use to profile, track and identify people and screen, sort and even nudge their behaviour can have a chilling effect on the freedom of expression and the freedom of assembly and association. Using facial recognition in public areas may interfere with a person's freedom of opinion and expression, simply because the protection of 'group anonymity' no longer exists. This could discourage people to attend demonstrations and join in peaceful assembly, which is one of the most

⁶ Vian Bakir, "Psychological Operations in Digital Political Campaigns: Assessing Cambridge Analytica's Psychographic Profiling and Targeting," *Frontiers in Communication*, Vol. 5 (2020). <https://www.frontiersin.org/article/10.3389/fcomm.2020.00067>

⁷ Jessica Fjeld et al., "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, 15 January 2020). <https://doi.org/10.2139/ssrn.3518482>

important elements of democratic society. Individuals may also prefer to refrain from expressing certain points of view and accessing some sources of information if they fear that the data collected on their activities may be used by AI technologies designed to take decisions on them (e.g., recruitment or promotion to a new position).

Diminishing Critical Thinking and Decision-making

AI technologies have been increasingly commissioned within shared public spaces to even make decisions and take actions for the sake of efficiency and speed. Despite these narrow gains, a better understanding of attitudes toward and interactions with AI technologies and underlying models is essential, precisely because of the aura of objectivity and infallibility cultures tend to ascribe to them.⁸ As a result, over the last decades, one has witnessed a certain degree of de-politicization of decision-making and reliance if quantitative model considered highly objective. The technologies have contributed to this.

This trend can result in passivity amongst citizens, rather than encouraging them to question and critically research the reasons for the choices made and to know such choices are rooted in interests or values which need not necessarily be unobjectionable, absolute, or “scientific” to be valid. Accustoming society to accepting choices not based on critical reasoning but according to the dictates of authority is extremely unjust, and therefore harmful, given that it is impossible to establish, in a provable manner, who should be regarded by public opinion as an authoritative source. AI-assisted technologies may make people believe that they are making their own choices, whereas they are merely following patterns. More broadly, AI-assisted political decision-making may ultimately lead to establishing a form of automated democracy and depriving humans of autonomy over political processes. Defining societal goals should not be left to algorithms and must remain with humans enjoying democratic legitimacy and assuming political and legal responsibility.

Information Asymmetry

AI technologies are fuelling a potent threat to democracy in form of an unprecedented and mostly un-checked concentration of data, information, and power in the hands of a small group of major digital companies which develop and own the algorithms, as well as the centralization of the Internet itself. These big companies no longer serve as simple channels of communication between individuals and institutions but play an increasingly prominent role on their own, setting the agenda and shaping and transforming social and political models. If too much political

⁸ Osonde A. Osoba and William IV Welsler, *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*, (RAND Corporation, 5 April 2017). https://www.rand.org/pubs/research_reports/RR1744.html

power is concentrated in a few private hands which prioritize shareholder value over the common good, this can threaten the authority of democratic states. Thus, there is a clear need to reduce the influence of major private companies on democratic decision-making. Moreover, public-private collaborations in AI and its use in sensitive fields, such as public order; security and intelligence; border control, but also in research and development, blur the boundaries between the responsibilities, processes and institutions of democratic states, and the interests of private corporations.⁹

Further, AI may also facilitate abuses of power by democratic states and state agencies: as a dual-usage technology, it can be deployed to undermine important human rights that are integral to the functioning of democracies. Advances in AI-based surveillance technology, such as facial, voice and motion recognition, together with a network of surveillance cameras in public places, allow the tracking of individuals in the real world. These AI capacities have come to the forefront during the Covid-19 pandemic. As with progress in other technologies, tools for surveillance together with predictive analytics can both be used to increase security, safety, or traffic control, as well as enable governments to control large crowds and predict the formation of protest and riots.¹⁰ Thus, AI-driven blanket surveillance measures threaten an individual's right to privacy and to freedom of expression.

To sum up the challenges, democracy implies that people with different views should be able to come together to find common solutions through dialogue.¹¹ Instead of creating a public common space and a common agenda, AI-based platforms seem to favour individualistic and polarised attitudes and lead to the emergence of closed homogenous virtual communities sharing the same views, thus undermining social cohesion, acknowledging real-world heterogeneity and facilitating robust democratic debate. In contrast, AI technologies continue contributing to proliferation of hate speech, compartmentation, and segmentation of society. The problem is exasperated by the fact that segments of the population are not even represented in these platforms due to various legacy inequities (i.e., based on gender, age, socio-economic status) also needs to be factored in this reflection. Private companies which apply the rules of the market and not those of democracy are taking no responsibility for allowing the fuelling hate speech and distributing violent content.

Without doubt, AI technologies provide powerful mechanisms to interfere with the

⁹ Council of Europe, (4 October 2019).

¹⁰ *Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications*, Council of Europe, 2017. <https://edoc.coe.int/en/Internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html>

¹¹ Noémi Bontridder and Yves Pouillet, "The Role of Artificial Intelligence in Disinformation," *Data & Policy*, Vol. 3 (ed 2021). <https://doi.org/10.1017/dap.2021.20>

processes of democracies and undermine democratic institutions. The use of AI, and its potential for abuse by governments, and by private corporations, poses a real threat to the institutions, processes, and norms of rights-based democracies.

Solution Approach

Tackling these challenges may seem daunting. So, the question arises what can be done when approaching these challenges to minimize the threats posed, while not stifling the potential of AI technologies. As AI systems can be complex and bear significant risks in certain contexts, building trust is essential. Clear rules need to address high-risk AI systems without putting too much burden on less risky ones to keep fostering innovation and leverage the true potential of AI technologies. The solution could include mobilizing resources, in partnership with the private and the public sector, along the entire value chain and to create the right incentives to accelerate deployment of AI, including by smaller and medium-sized enterprises. This includes working with member states and the research community, to advance positive developments of AI.

There is an obvious gap between the pace of technological development and the regulatory framework. Self-regulatory principles and policies cannot be the only tools to regulate AI as they do not lead to accountability. We need to ensure that the power of AI is regulated and used for ‘common good’, with specific principles encoded based on the protection of human rights, democracy, and rule of law. The contextual assessment needs to be a multi-stakeholder process, including citizens, government, civic societies, and private companies. To ensure accountability, the legal framework should be put in place should provide for independent oversight mechanisms that would guarantee effective compliance with its provisions.¹²

However, such an oversight mechanism can only be effective if it can be proactive and engaged *ex ante*. Indeed, while it would be important to introduce sanctions for non-compliant behaviour, a mechanism that would limit itself to *ex-post* penalties and fines - which are usually easily affordable by big private companies no matter the amount - would not achieve the desired outcome. That is because it is often very difficult, if not impossible, to restore the previous situation or “erase the damage” after a given AI technology has been introduced and used, as unethical and/or non-compliant with human rights, democracy, and rule of law as it may be.

A proactive oversight mechanism requires a highly competent (in technical, legal, and ethical terms), capable of following the new developments on digital technology and

¹² PACE - *Doc. 14868* (2019) – “Need for Democratic Governance of Artificial Intelligence,” accessed on 16 February 2022. <http://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=27616&lang=en>

evaluating accurately and authoritatively its risks and consequences.¹³ More critically, the role of AI in changing the power balance between institutions, political actors, and executive organs needs more structured research and tighter loop with policymaking. Given the scale of legitimacy and sovereignty problems relating to outsourcing political decisions to algorithms, the role of constitutions, parliaments, and the political elites in relation to AI needs to be studied in-depth with a specific focus on how political authority should be situated in the age of automated decisions.¹⁴

This does not mean that AI cannot be a force for good, or render politics more efficient, or more responsive to citizens' needs. If used well, AI can broaden the space for democratic representation by decentralizing information systems and communication platforms. It can bolster informational autonomy for citizens and improve the way they collect information about political processes and help them participate in these processes remotely. Just as AI can be used to strengthen opacity and unaccountability, it can also improve transparency and help establish greater trust between the state and the society and between citizens themselves.¹⁵

In line with the outlined solution approach, UNESCO's Recommendation on the Ethics of Artificial Intelligence, includes not only ethical principles and values but also concrete policy actions along with defined tools like Readiness Assessment and Ethical Impact Assessment.¹⁶ The recommendation includes specific provisions like adding the role of an independent '*AI Ethics Officer*' along with other tangible mechanisms to oversee ethical impact assessment, auditing and continuous monitoring efforts and ensure ethical guidance of AI technologies in the public domain. It includes guidelines around provisioning of redressal mechanisms for citizens to help bring agency to the impacted while laying stress on inclusiveness, gender equality, trustworthiness, the protection of environment and privacy. The Recommendation is intended to serve as a global ethical benchmark emphasizing on contextual assessments and equitable governance models. For its part, UNESCO, as a leading international standard-setting organization in the field of societal transformation, aims to play a pioneering role in designing ways and formats to ensure that AI technologies are used to enhance democracy through citizens' deliberative and participatory forms of people's involvement in democratic processes.

¹³ Stephanie Weiser, "Requirements of Trustworthy AI," *FUTURIUM* - European Commission, (8 April 2019). <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1>

¹⁴ Hamid Akin Ünver, *Artificial Intelligence, Authoritarianism, and the Future of Political Systems*, EDAM, (2018). <https://doi.org/10.13140/RG.2.2.19598.00329>

¹⁵ *WHITE PAPER On Artificial Intelligence - A European Approach to Excellence and Trust* (2020). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:65:FIN>

¹⁶ "Recommendation on the Ethics of Artificial Intelligence," UNESCO, (27 February 2020), <https://en.unesco.org/artificial-intelligence/ethics>