DEMOCRACY IN THE AGE OF AI; THE FINE LINE BETWEEN THE KNOWN AND UNKNOWN

Mihai SEBE, Alexandru GEORGESCU and Eliza VAȘ[*]

Declaration[*]

We are currently witnessing a structural shift in the democratic societies across the globe that impacts the internal structure of politics. The rapid development of Artificial Intelligence (AI) and the increased level of cyber security threats affect political socialisation in ways that could be observed before in a generation's time, not just in a matter of years as it is the case now. Therefore, we need to start thinking about the rhetorical shift of digitalisation towards AI, given its applicability as tool for political deliberation, preference aggregation and discovery, as well as electoral integrity. The technological advancement can impact social stability, all while not offering antibodies the democratic societies need to deal with this disruption.

The future of politics seems to be more defined by the way in which states will protect the minds, the will and the hearts of their citizens against cognitive warfare and threats. On a decision-making level, this further burdens the process of making the AI an instrument for authentic public participation. Between informing choices and decisions and delegating free will to AI lies a fine line. As such, a new social contract is needed.

In a world dominated by geopolitical tensions, the multilateral approach offers us ideas and ideational tools to draft this new contract. In this context, the United Nations represent a global standard that one cannot ignore.

The malicious use of digital tools is high on the international and national agenda and is set to remain a pressing issue for the upcoming years. In fact, the United Nations went one step

[*] Mihai SEBE, PhD, lecturer, University of Bucharest. E-mail: mihai.sebe@fspub.unibuc.ro,
Alexandru GEORGESCU, PhD, is an Established Researcher (R3) with the National Institute for Research and Development in Informatics ICI Bucharest, alexandru.georgescu@ici.ro,
Eliza VAȘ is an expert in European affairs and vice-president of the Young Initiative Association. E-mail: eliza.vas@younginitiative.org.

[*] The opinions expressed belong to the authors only and do not reflect the official position of the institutions they are affiliated with.

further and signalled in 2024 the impact on politics. "Electoral security posed a significant challenge throughout the year, as elections took place in more than 70 states. Disinformation campaigns designed to influence voters, and uses of deepfakes of political figures were observed in the lead-up to national elections in several states. At the same time, interference in critical infrastructure facilitating electoral processes also posed risks." (United Nations, 2024).

Following this assessment, the UN General Assembly adopted the Pact for the Future (*Resolution 79/1*), including the Global Digital Compact (*Annex I*). It calls for closer international cooperation that reduces all digital divides between and within countries and reiterates that the digital future should be guided by the purposes and principles of the Charter of the United Nations (United Nations General Assembly [UNGA], 2024). Thus, the Human Rights Council chose to anchor the new developments in the framework of the international law stating explicitly this duality: the new and emerging digital technologies can hold great potential for strengthening democratic institutions and the resilience of civil society, but they can also affect the integrity of democratic institutions (*Resolution 59/11, 2025*) (United Nations Human Rights Council [UNHRC], 2025a).

We are now in a situation where the quick evolution of AI needs to meet a series of standards, such as a human rights-based approach and appropriate safeguards and human oversight. Moreover, we are faced with an unprecedented spread of disinformation the AI can only make worse (United Nations Human Rights Council [UNHRC], 2025b).

As a technology, AI comes with a paradigm shift from a security perspective. The cyber security we are used to must make way for new approaches, especially when it comes to generative AI embodied in the vast majority of systems with which average citizens will interact in a media, social and political context. On the one hand, we rely on this technology for AI aggregators of information, AI-based communication, AI pollsters and AI-based societal systems (education, public services etc.). However, AI diverges from traditional software by utilising natural language interfaces and generating probabilistic rather than deterministic outputs. This lack of a predictable baseline makes it difficult to detect anomalies.

Furthermore, because AI resilience is deeply tied to specific data environments and usage patterns, conventional security audits often prove inadequate. Experts are recommending redteaming, which is a structured ethical adversarial testing under real conditions (United

Nations Educational, Scientific and Cultural Organization [UNESCO], 2025). However, there is a lack of specialists, especially within user entities as opposed to AI development entities, leading to a “black box” problem where AI tools become opaque and anomalous behaviours become harder to identify. On the other hand, we have AI systems as tools for cyber attacks on electoral infrastructure (some of which may also be AI-based) or through generation of fake news, of deepfakes, of tailored disinformation and through the manufacturing of public moods and ideas. This is an incredibly pernicious issue. The identification of AI-generated content before destabilising a narrative or an entire society is a difficult task, likely involving AI tools, new skillsets, new modes of societal protection and new tools used ethically under an acceptable and values-based legal and administrative framework.

This complicated relationship between AI and democracy has been further explored in a UNESCO report that underlined both the great expectations and fears of this process, underlying that “the only political certainty we have today is that politics in the future will inevitably be very different from politics in the past”. The report warns of the erosion of public discourse, the rise of new intermediaries, and the opacity of algorithmic decision-making. This should be coupled with secular trends involving the shrinking of party membership as principal means of political activity, education and legitimation, as well as the erosion of the authority of public institutions and official narratives in a digitalized society. The digital utopianism needs to make room for more balanced democratic nuances as the AI tends to reflect the values of its creators and the biases in its datasets instead of remaining neutral. The evolution of AI raises two main questions as per this report: “Do the principles of democratic self-governance still hold relevance and significance in a digital, automated public space largely governed by algorithmic systems? Is this a new era that we must simply accept, or does this historical moment bring new opportunities for democratisation?” (Innerarity, 2024).

While decision-makers all around the world are looking for answers to these questions, the European Union (EU) is proposing a risk-based approach. With the AI Act introduced in 2024, the EU mapped out four levels of risk for AI systems. From minimal to unacceptable risk, the purpose was to make sure that AI brings more benefits to the citizens than it takes away freedoms and rights. Under high-risk category we have those situations that can severely affect fundamental rights, safety and health. With respect to democracy, the Regulation foresees that “AI systems intended to be used to influence the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote in elections or referenda should be classified as high-risk AI systems with the exception of AI systems whose output natural persons are not directly exposed to" (Regulation (EU) 2024/1689 [AI Act], 2024).

Coming back to the will of the people as a core democratic principle to uphold, we invite the readers to reflect on the words of Stephen Hawking, who 10 years ago affirmed that "in the future, AI could develop a will of its own — a will that is in conflict with ours" (University of Cambridge, 2016). Furthermore, in seeking to develop national and international governance frameworks and accepted principles of AI development, we should be way of the differing rates of digitalization throughout various societies. As the famous science fiction author William Gibson once observed, "the future is already here, but it is not evenly distributed" (O'Toole, 2012).

References

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L 2024/1689. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689

Innerarity, D. (2024). *Artificial intelligence and democracy*. Paris, France: UNESCO. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000389736

O'Toole, G. (2012, January 24). *The future has arrived — It's just not evenly distributed yet*. Retrieved from http://quoteinvestigator.com/2012/01/24/future-has-arrived/

UNESCO. (2025). *Red teaming artificial intelligence for social good: The playbook*. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000394338.locale=en

United Nations General Assembly. (2024). *Resolution adopted by the General Assembly on 22 September 2024* (A/RES/79/1). Retrieved from https://docs.un.org/en/A/res/79/1

United Nations Human Rights Council. (2025a). *New and emerging digital technologies and human rights* (A/HRC/RES/59/11). Retrieved from https://digitallibrary.un.org/record/4087001?v=pdf

United Nations Human Rights Council. (2025b). *New and emerging digital technologies and human rights* (A/HRC/59/L.14). Retrieved from https://docs.un.org/en/a/hrc/59/l.14

United Nations. (2024). *United Nations disarmament yearbook: Developments and trends 2024*. Retrieved from https://yearbook.unoda.org/en-us/2024/chapter5/

University of Cambridge. (2016, October 19). *"The best or worst thing to happen to humanity": Stephen Hawking launches Centre for the Future of Intelligence*. Retrieved from https://www.cam.ac.uk/research/news/the-best-or-worst-thing-to-happen-to-humanity-stephen-hawking-launches-centre-for-the-future-of