AI CHALLENGING SOVEREIGNTY AND DEMOCRACY

AI is wonderful. AI is scary. AI is the path to paradise. AI is the path to hell. What do we make of these contradictory images when, in a world of AI, we seek to both protect sovereignty and respect democratic values? Neither a techno-utopian nor a dystopian view of AI is helpful. The direction of travel must be global guidance and national or regional AI law that stresses end-to-end accountability and AI transparency, while recognizing practical and fundamental limits. We also need a deeper understanding of tensions between AI and sovereignty and democratic values, in order to address them in the interplay of the social construction of sovereignty and democracy and the technological construction of AI.

Paul Timmers*

TPQ
Winter 2021/22

* Prof Dr Paul Timmers is research associate at the University of Oxford (Oxford Internet Institute), professor at European University Cyprus, visiting professor at KU Leuven and University of Rijeka, senior advisor to EPC Brussels, board member of Digital Enlightenment Forum and President of the Supervisory Board of the Estonian eGovernance Academy.

AI is wonderful. AI is scary. AI is the path to paradise. AI, it is the path to hell.

We read about the fantastic progress and achievements of AI. It is more accurate than doctors in diagnosing eye disease.[1] Able to discover new medicines.[2] Writing software and stories.[3] AI helping to provide personalized education so that our children can find good jobs in the future. AI to weed out fraud in public services, and thereby shore up the legitimacy of government.

But we also read about AI which is used to spy upon citizens. So-called to protect the state but in fact having a chilling effect on freedom of speech and association. AI is enabling facial recognition for public surveillance to suppress whole populations. AI which creates fake online persona in images and words that fool us thinking that an impersonated authority really said something that they actually never said, undermining democracy.

What do we make of these contradictory images? Can we reconcile AI with both protecting sovereignty and respecting democratic values?

In any event, neither a techno-utopian nor a dystopian view of AI is helpful. We must evolve our understanding for a balanced view and avoid rigid thinking. We need to develop an AI policy which unlocks AI innovation and benefits. We must seek protection and progress in the interplay of the social construction of sovereignty and democracy and the technological construction of AI.

Sovereignty and Democracy

We want to understand what AI means for sovereignty. State sovereignty is a political concept, that is not strictly defined and sometimes confusing. It concerns territory and borders, natural and digital resources 'that belong to us', people, and authority that has internal and external legitimacy. Internal legitimacy refers to the effectiveness of the state for the citizens, such as delivering well-functioning public services and fair justice. Internal legitimacy is also about recognition by citizens of the government, such as confidence in the rule of law. External legitimacy concerns the recognition by foreign states and the autonomy of a state towards third states.[4]

[1] Jeffrey de Fauw, Joseph R. Ledsam, Bernardino Romera-Paredes, Stanislav Nikolov, Nenad Tomasev, Sam Blackwell and Harry Askham, "Clinically Applicable Deep Learning for Diagnosis and Referral in Retinal Disease." *Nature Medicine*, Vol. 24, No. 9 (2018): p. 1342–50. https://doi.org/10.1038/S41591-018-0107-6

[2] World Economic Forum, "How Is Artificial Intelligence Being Used in Medicine? | World Economic Forum," (2021). https://www.weforum.org/agenda/2021/07/ai-discover-new-drugs-nature/

[3] AlphaCode Team, "Competitive Programming with AlphaCode | DeepMind," (2022). https://deepmind.com/blog/article/Competitive-programming-with-AlphaCode

[4] Lokke Moerel and Paul Timmers, "Reflections on Digital Sovereignty: EU Cyber Direct." *Research in Focus*, (January 2021). https://eucyberdirect.eu/research/reflections-on-digital-sovereignty

AI can both strengthen and undermine these forms of legitimacy.

We also want to understand what AI means for democracy and democratic values. For these too we have plenty of reference points, such as democracy and democratic culture as defined by the Council of Europe or as codified in EU law.[5] Recently a European Declaration on Digital Rights and Principles for the Digital Decade was proposed.[6] This stresses general values that should also hold in the digital world, such as people at the centre, solidarity and inclusion. It also insists on specific values that can be at stake with AI, such as freedom of choice, transparency, fair treatment, freedom of expression and information, and protection against disinformation.

"In all three cases legitimacy of the state in delivering public services - an important pillar of sovereignty - got undermined as an unintended consequence of the use of AI."

Finally, we know well that sovereignty and democratic values can reinforce each other, but can also be at odds. Some even argue that in the global economy there is a battle between sovereignty and democracy.[7] AI can be both positive and negative in the relation between sovereignty and democracy.

Public Services

The cases mentioned before have outcomes that are intentional, that is, they deliver what the AI designers intended. But increasingly we become aware of unintended consequences of AI. Let's discuss some of those.

In Austria, an AI-based labour market opportunities system was set up. The AI program assesses the probability of integration into the labour market. The system was seen by the government as a promise to combat discrimination against women, disabled and elderly persons. But scientists found the system simplifies the reality of individuals into a single number. They warned that personalized advice on jobs and work would degenerate into a generalized approach, that doesn't do justice to individuals. They requested the system to be put on hold.

[5] European Union, "Charter of Fundamental Rights of the European Union," *Official Journal*, OJ C, No. 326 (2012): p. 391–407. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT

[6] European Commission, "Declaration on European Digital Rights and Principles | Shaping Europe's Digital Future." (2022). https://digital-strategy.ec.europa.eu/en/library/declaration-european-digital-rights-and-principles

[7] Jonathan Derbyshire, "Why Governments Can't Have It All | Financial Times." *Financial Times*, 28 July 2017. https://www.ft.com/content/63246e18-72b4-11e7-aca6-c6bd07df1a3c

In the UK, a machine learning system was piloted to identify which children could be at risk and whether such risk could escalate. Tens of local authorities already started using the system. However, according to What Works for Children's Social Care (WWCSC), the models were wrong six times out of ten. WWCSC called to first ensure more transparency of experiences.

From 2011 onwards the Dutch tax authority had been using a self-learning risk-classification model to detect fraud in claiming child benefits. This touched 26,000 parents and 70,000 children. The political mood at the time was hardening due to an earlier social benefits fraud case involving Bulgarians. Combatting child benefits fraud became a political priority. In a hard-nosed, business-like approach, parents had to pay back full benefits. Some parents lost jobs and homes. Children got placed out of their families. Unjustified in a number of cases or based on dubious suspicions, to say the least.

Namely, unfortunately, the AI embedded discrimination in its risk-classification model, based on nationality. On the resulting injustice, despite several inquiries, little corrective action was taken. A parliamentary report talked of 'tunnel vision' and 'no one taking responsibility'. Ultimately, in 2021 the Dutch government fell over the scandal. The Netherlands showed that it is a democratic state which upholds accountability and transparency. Nevertheless, trust in government got a severe hit: 71 percent of people felt government legitimacy got eroded.[8]

In all three cases legitimacy of the state in delivering public services - an important pillar of sovereignty - got undermined as an unintended consequence of the use of AI. And there are many more such cases, all over the world!

Public Security - Facial Recognition and Surveillance

AI performs impressively on facial recognition. A whole industry has sprung up, supplied by observation camera manufacturers mostly from China. Cheap cameras are now installed by the millions in both democratic and authoritarian states. The industry also consists of facial data and recognition companies such as ClearView, which has built up a huge image database. Facial recognition has become a popular technology in public surveillance.

On the one hand, police and judicial authorities stress the urgent need for facial recognition to track criminals and terrorists. These voices speak for the public interest and argue the need to defend the state, i.e., state sovereignty, for instance

[8] I&O Research, *Vertrouwen in Overheid Na Drie Grote Kwesties Uitgevoerd in Opdracht van NRC Handelsblad*, (2021).

against terrorism. However, facial recognition is also an instrument of control of techno-authoritarian states (as in China for the suppression of the Uighur population).

The EU has put forward an AI Act which defines four risk classes for AI: unacceptable, high risk, limited risk and minimal risk[9]. The AI Act puts the use of facial recognition in public spaces for law enforcement purposes in the unacceptable risk category. The law bans, subject to narrow exceptions, live remote biometric identification systems in publicly accessible spaces used for law enforcement purposes. The narrow exceptions (with further specific constraints) relate to crime victims, terrorist attacks, and serious criminals. At the time of writing, the AI Act is being negotiated between the European Parliament (EP) and the Council of Ministers (i.e., the EU Member States).

> *"Only with the help of AI can we timely recognise an emerging attack and monitor the thousands of hardware and software elements of a critical infrastructure. We need AI to adapt firewalls, stop an attack from spreading and counter malware mutations."*

In the meantime, the EP has adopted a non-binding Resolution that states: "conducting facial recognition in public places, as well as at automatic border control gates used for border checks at airports, may pose specific risks to fundamental rights."[10] The EP called for a moratorium on the use of facial recognition technology in public places, and a ban on any processing of facial images for law enforcement purposes that leads to mass surveillance in publicly accessible spaces. It also wants a ban on the use of private facial recognition databases in law enforcement (referring specifically to the use of the maligned ClearView).[11] Data protection authorities in the EU go even further and call for a general ban on any use of AI for automated recognition of human features in publicly accessible spaces.[12]

Generally, the fear is that such surveillance threatens a range of democratic values such as the right to fair justice, and freedom of expression and association without

[9] European Commission, "Proposal for a Regulation on Artificial Intelligence, COM/2021/206 Final," (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206

[10] European Parliament, "Resolution on Artificial Intelligence in Criminal Law and Its Use by the Police and Judicial Authorities in Criminal Matters," (2021).

[11] European Parliament, (2021), pt. 28.

[12] EDPB and EDPS, "EDPB-EDPS Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act)," (2021).

fear of being censored or intimidated. Still, there can be a justified need to defend the state with the help of such AI technology. The EU AI Act proposes to strike a balance, but clearly not everyone agrees.

One may wonder, though, whether we are fully and sufficiently flexibly combining law and technology for such use of AI? Could there be a better interplay between law-based governance (risk assessment, oversight, accountable exemptions, standards, auditing, conditions for access to markets) and technology requirements to evolve AI software and hardware such as cameras (transparency, replaceability, upgrading, data deletion – right to be forgotten, all embedded in software and hardware)? Surely, this would mean that the world of law-makers and the world of tech-makers need to work more closely hand-in-hand. In order to avoid regulatory capture and to safeguard both democratic values and sovereignty, such collaboration should be subject to democratic oversight.

Norms and Values

The legitimacy of authorities and government can be at risk when AI is deployed without considering the larger societal system of public service delivery, judiciary, and political accountability. AI itself also embeds values, explicitly or not, as illustrated by the Austrian and Dutch cases. These may clash with democratic values and can lead to discrimination on socio-demographic parameters.

AI can also be a powerful ideology machine.[13] For instance, to promote dataism – the ideology of replacing the person by their data representation. Yuval Harari states that "we may interpret the entire human species as a single data processing system, with individual humans serving as its chips".[14] AI can also provoke a counter-ideology, inciting a Luddite human vs machine confrontation. Alternatively, AI tempts some to adhere to an omnipotent-AI ideology or revive old-fashioned We-Can-Be-In-Control ideology.[15]

AI creates new possibilities and new opportunities. To choose from these, new norms are needed. What then are these norms? Above all, who determines the new norms and their underlying values? Are they decided by the large platform companies that invest most in AI (in 79 percent of all papers on AI, authors now

[13] Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, (Yale University Press: 2021). https://bookshop.org/books/atlas-of-ai-power-politics-and-the-planetary-costs-of-artificial-intelligence/9780300209570

[14] Yuval N. Harari, *Homo Deus: A Brief History of Tomorrow*, (Harper, 2017).

[15] An ideology that got a severe hit with the COVID-19 pandemic. Helga Nowotny, *In AI We Trust: Power, Illusion and Control of Predictive Algorithms* (Polity, 2021).

have corporate ties)?[16] Such corporate giants know well how to get such norms politically and legally supported in the processes of governmentalism.[17] Are corporate values reflecting the ethics that would result from democratic debate? These are important questions to answer for the future of ethical AI, the future of democratic values, and indeed, the future of sovereignty.

The debate on ethical AI has already resulted in nearly 100 guidelines, that widely diverge, but also have as common themes transparency, fairness and justice, non-maleficence, legal responsibility, and privacy.[18] These guidelines are, however, mostly originating in Europe and North America, while there is a marked absence of guidelines from the Global South. With the help of comparative studies, we may, however, still achieve global guidance on ethical AI.[19]

Transparency

One magic wand to tackle the conundrums of AI is 'transparency'. However, AI is a complex socio-technical system of data, hardware/software architectures like cloud, AI algorithms, decision systems, in combination with the skills and mindset of designers, coders, commercial providers, civil servants, etc. The parts relate to and influence each other. AI is a system technology[20]. It is not so easy to create end-to-end transparency in such an AI system. Already it is very hard to explain how many deep learning algorithms get to their conclusions. There are also difficult questions about the data that are used to train algorithms. How do we know that they are not biased and not been tampered with, are they still available after the algorithmic training has been concluded, can personal data be kept confidential, etc. But uncovering the socially constructed part of the AI system is hard too. Transparency is easier said than achieved.

Fortunately, some scientists and developers are willing to take up these questions

[16] Abeba Birhane, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan and Michelle Bao, "The Values Encoded in Machine Learning Research," (June 2021). https://arxiv.org/abs/2106.15590v1

[17] *Julie E. Cohen, Between Truth and Power: The Legal Constructions of Informational Capitalism.* (Oxford University Press, 2019).

[18] Antonio Casilli, "What Is a 'Truly Ethical' Artificial Intelligence? - YouTube." *Digital Humanism Lectures*, (25 January 2022). https://www.youtube.com/watch?v=9NWSgny12wY; Mariarosaria Taddeo and Luciano Floridi, "How AI Can Be a Force for Good – An Ethical Framework to Harness the Potential of AI While Keeping Humans in Control." *Philosophical Studies Series*, Vol. 144 (2021): p. 91–96. https://doi.org/10.1007/978-3-030-81907-1_7

[19] Luciano Floridi, Josh Cowls, Thomas C. King and Mariarosaria Taddeo, "How to Design AI for Social Good: Seven Essential Factors." *Science and Engineering Ethics*, Vol. 26, No. 3 (2020): p. 1771–96. https://doi.org/10.1007/S11948-020-00213-5/TABLES/1; Michael Dukakis, Nguyen Tuan and Marc Rotenberg, "Artificial Intelligence and Democratic Values - 2020 - CAIPD." (2020). https://www.caidp.org/reports/aidv-2020/

[20] Corien Prins, Haroon Sheikh, Erik Schrijvers, Eline de Jong, Monique Steijns and Mark Bovens, "Mission AI. The New System Technology | Report | The Netherlands Scientific Council for Government Policy," (2021). https://english.wrr.nl/publications/reports/2021/11/11/summary-mission-ai

and design new techniques that meet end-to-end requirements for AI to respect democracy and fundamental values. Technically, for instance, one part of the answer is to analyse confidential data without ever expositing them with the help of homomorphic encryption and multi-party computation. Design templates and value-based design are techniques that can help to embed values into technology. Corporate responsibility with standardized auditing and reporting can help too.[21] The lesson is clear: we do not need to accept that AI technology is a given and driven purely by commercial motivations, we can insist on getting end-to-end transparency in AI.

Profound Challenges

The wand of transparency may not suffice, though. Stuart Russell gives an example of a commercial algorithm designed to maximize click-through and thus generate more revenue. It is not that these algorithms then result in presenting information to the user that they like. Rather, the user's preferences are changed so that the users become more predictable. Who tends to be more predictable? Right, users with extreme political views. So, these algorithms aid the rise of anti-democratic thinking and group formation – challenging democratic values for sure, and perhaps also sovereignty.[22]

Here we have a consequence that is not readily undone by simply insisting on responsible and transparent AI. It gets worse when AI is intentionally deployed for bad intent, when it gets weaponized in social media information warfare. A dystopian future that is near is that your president is spoofed with the help of an AI-generated video and citizens are made to believe that an attack is imminent. Even law, that allows to take down maleficent AI based on post-market auditing (in analogy to what is called post-market surveillance of medicines) would then be inadequate. By the time the problem has been spotted, it may be too late: the Capitol has been stormed and taken.[23]

AI, we love it and hate it. We cannot live without it, but fear when we have to resort to it. Society is ever more complex and fast moving. Stock markets trade in milliseconds, news spreads in seconds, a pandemic explodes in a matter of weeks, the speed of adoption of new technologies has shrunk over a century from 30 years to 5 years. Such complexity and speed surpass human capability. We need AI to deal with this new world. However, doing so we wander into unknown territory.

---

[21] Jakob Mökander, Maria Axente, Federico Casolari, Luciano Floridi and J Mökander, "Conformity Assessments and Post-Market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation," *Minds and Machines* (2021). https://doi.org/10.1007/s11023-021-09577-4

[22] Stuart J. (Stuart Jonathan) Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (Penguin Books, 2020), p. 8.

[23] A memetic reference to 6 January 2021.

AI and Cybersecurity – an Intractable Challenge?

Let's consider AI and cyber resilience. Cyber resilience is about keeping our critical infrastructures such as electricity, water supply or hospitals running in the face of cyber incidents affecting the underpinning digital networks and systems. Increasingly, we cannot do without AI in the defence against cyber-attacks. The reason is the complexity of critical infrastructures (think of all the parts in an electricity network) and the speed of the digital world. Cyber-attacks are moving very fast, with computer viruses spreading in milliseconds. Malware may even be using AI to evade defences, and exploit vulnerabilities, while adapting itself in milliseconds. It can bring down a critical utility, even a whole country in a matter of minutes. Harbingers are the Sandworm attack on the electricity network in Ukraine in 2015, the NotPetya malware in logistics in 2017, and the DarkSide ransomware in the Colonial pipeline in the USA in 2021.

Only with the help of AI can we timely recognize an emerging attack and monitor the thousands of hardware and software elements of a critical infrastructure. We need AI to adapt firewalls, stop an attack from spreading and counter malware mutations. To contain an attack a switch-off may be needed as part of a critical service. Perhaps shutting down part of the electricity network and doing so in a matter of seconds or less. At machine timescales rather than human timescales. We cannot but leave this critical decision to AI even if the consequences are about humans. What do we switch off? Which city? Which industry? Which hospital?

Properly ensuring cyber-resilience is a legal obligation for critical services and infrastructures such as specified by the EU's Network and Information Security Directive.[24] But laws are based on a world of humans. For instance, they assume that experts convene when there is a cyber-attack. Such governance is a good way to ensure proper decision-making and accountability amongst humans. However, it is a way of working that is inadequate for the machine reality of cyber-attacks. The country may have come to a standstill by the time an expert picks up the phone.

So, we have no choice but to leave the decision to AI. But then, who is accountable? What if people die on their way to hospital as a consequence of the AI switching off an electricity system? Could it have been avoided? Who chooses who will die and who will be spared? What if the AI intervention is endangering the state itself?

Law operationalises internal legitimacy of the state. It is an instrument of sovereignty and a sovereign instrument. Democratic societies have political accountability for

[24] European Union, "Directive (EU) 2016/1148 Concerning Measures for a High Common Level of Security of Network and Information Systems across the Union," *Official Journal*, OJ L, No. 194 (2016): p. 1–30.

the use of the law. But how do we deal here with this? Which minister or politician can be held accountable, if any? Therefore, indispensable AI (in this case of cybersecurity) creates a fundamental challenge for sovereignty and for democracy.

In extremis, perhaps even of an existential nature[25] this challenge poses itself when AI is put into kinetic weapons. These lethal autonomous weapons (LAWS) make decisions without human intervention. Examples are autonomous drones with target recognition and hypersonic autonomous missiles. These machines decide if and where to hit. They decide on life and death. Might they provoke war where human diplomacy could still have run its course?

Ways Forward

What then can we do to safeguard democratic values and safeguard legitimate sovereignty in a world of AI? First, we need to recognize the challenges and create visibility about what is going on, how AI can be a threat and how it is actually deployed and what we can expect to come. We need to stress end-to-end accountability and AI transparency, while recognizing their practical and possibly fundamental limits.

Secondly, the challenge goes beyond a single unethical or ethically-unaware AI developer, a single social media company, a single cybersecurity agency, and beyond a single government. For computability of AI, sovereignty and democratic values we need to put in place a whole-of-society approach. Even if that is hard to do, this must be the direction of travel. Moreover, AI in many ways is both global and local in terms of its opportunities and in terms of its relationship with values and ethics. The international dimension needs to be recognized, as a problem and also as a place to find solutions. International agreement on the relation between AI and sovereignty will be very hard, but that should not deter efforts to collaborate on constraining AI from the common evil (e.g., autonomous weapons) and on promoting AI for the common good (e.g., public health and saving the planet). The overall framing then must be a set of rules, whether under international law, or established at regional or national level but taking into account the international context.

Thirdly, we need to deepen our understanding of the fundamental tensions between AI and sovereignty and democratic values. There are no easy answers as this concerns the ill-understood relation of social and technological constructs, the relationship 'code ⇔ law'. Perhaps we need transparent supervisory AI that controls operational AI and that works at human time scales. Perhaps we need a more

[25] Nick Bostrom, "The Vulnerable World Hypothesis," *Global Policy*, Vol. 10, No. 4 (2019), p. 467. https://doi.org/10.1111/1758-5899.12718

flexible law that foresees adaptation within the law of scope and governance. This may seem of academic interest mostly, but this is not true: with this we are at the heart of constructing sovereignty and democracy in the age of AI.[26]

[26] Paul Timmers, "The Technological Construction of Sovereignty." in *Perspectives on Digital Humanism* (Springer Cham, 2022), p. 213–18. https://doi.org/10.1007/978-3-030-86144-5_28