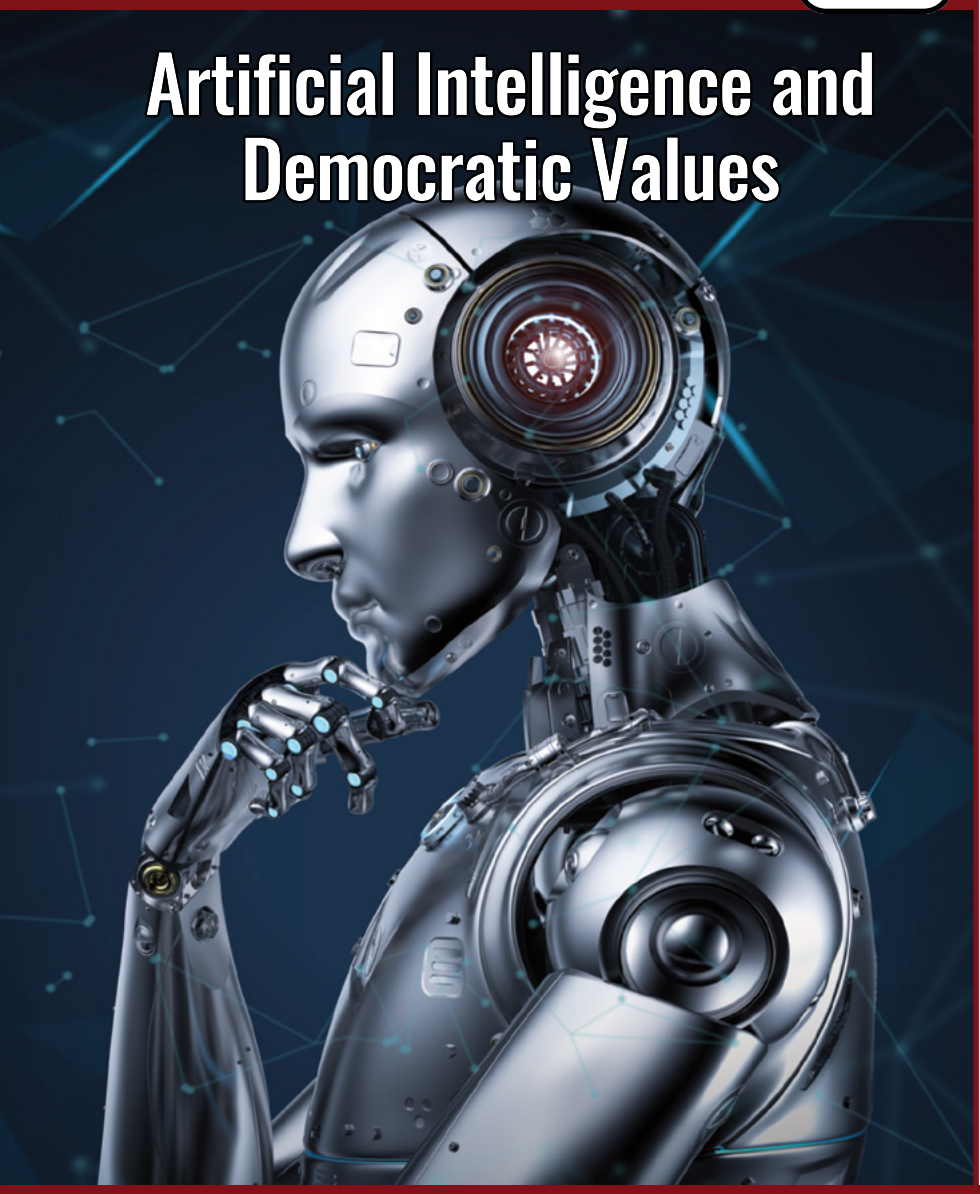
TPQ

• Winter 2021/22 • Vol. 20 • No. 4

20 YEARS

Artificial Intelligence and Democratic Values



80th ISSUE

HER BUSINESS, HER POWER!

We support female entrepreneurs with **Beko 100 Women Dealers Project** and encourage them to join the Beko dealer ecosystem with unique benefits.

Apply Now **beko.com.tr/100kadinbayi**

TPQ

Vol. 20 No. 4 / WINTER 2021/22

www.turkishpolicy.com

E-mail for Subscription: **subscriptions@turkishpolicy.com**
E-mail to the Editor: **editor@turkishpolicy.com**

Yayıncı **Toplumsal Katılım ve Gelişim Vakfı** Sorumlu Yazı İşleri Müdürü **M. Selim Alan**
Tüzel Kişi **A. Kemal Köprülü**
Yayın Türü **Yaygın Süreli Dergi** Görsel Yönetmen **Osman Balaban**
Baskı Organizasyonu **Özgür Matbaa Ltd. Şti.**
Litros Yolu Sok. Maltepe Mah. 2. Matbaacılar Sitesi Kat.3 IND4 Topkapı/Zeytinburnu /İstanbul
T: (0 212) 612 1360
info@ozgurmatbaacilik.com

TURKISH POLICY QUARTERLY
Asmalı Mescit Mah. Oteller Sok. No.8 Kat.2 Beyoğlu- İSTANBUL
T: (0 212) 621 4442 - 621 9258 **F:** (0 212) 531 8718
Web: www.turkishpolicy.com Email: info@turkishpolicy.com
Issue cover visual by Shutterstock

TABLE OF CONTENTS

THE PREMIUM CORPORATE SPONSOR OF THIS ISSUE:

Foreword

Then to Now

How much time is twenty years? Long enough to inspire, or short enough to be unnoticeable? Turkish Policy Quarterly (TPQ) was published for the first time in February 2002. We are celebrating its 20th anniversary with this issue. While much has changed since then, we believe the values that guide TPQ are as relevant and important as ever. There was then and there is now a chance for us all to contribute to a better world. TPQ has always adhered to and will continue to adhere to this ideal. In return, its global audiences have relied on TPQ since the very first day to provide them with credible, balanced, inter-disciplinary, and independent coverage. TPQ's evolution has been complex, involving constant efforts to understand new technologies and redefine fundamental concepts within the realm of policy. Over the last two decades, TPQ faced numerous challenges as the world experienced devastating depressions, wars, and economic and cultural changes as it became widely regarded as an influential journal. This really had nothing to do with luck. TPQ team and our great contributing writers have always been accountable for making TPQ's stories come alive. Yet, the question remains as to how long, or short, twenty years actually is.

It wasn't difficult to choose the focus of TPQ's 80th issue. Artificial intelligence has fast become part of everyday life, and we wanted to understand how it fits into democratic values. It was important for us to ask how we can ensure that AI and digital policies will promote broad social inclusion, which relies on fundamental rights, democratic institutions, and the rule of law. There seems to be no shortage of principles and concepts that support the fair and responsible use of AI systems, yet it's difficult to determine how to efficiently manage or deploy those systems today.

Merve Hickok and Marc Rotenberg, two TPQ Advisory Board members, wrote the lead article for this issue. In a world where data means power, vast amounts of data are collected every day by both private companies and government agencies, which then use this data to fuel complex systems for automated decision-making now broadly described as "Artificial Intelligence." Activities managed with these AI systems range from policing to military, to access to public services and resources such as benefits, education, and employment. The expected benefits from having national talent, capacity, and capabilities to develop and deploy these systems also drive a lot of national governments to prioritize AI and digital policies. A crucial question for policymakers is how to reap the benefits while reducing the negative impacts of these sociotechnical systems on society.

Gabriela Ramos, Assistant Director-General for Social and Human Sciences of UNESCO, has written an article entitled "Ethics of AI and Democracy: UNESCO's Recommendation's Insights. In her article, she discusses how artificial intelligence (AI) can affect democracy. The article discusses the ways in which Artificial Intelligence is affecting democratic processes, democratic values, and the political and social behavior of citizens. The article notes that the use of artificial intelligence, and its potential abuse by some government entities, as well as by big private corporations, poses a serious threat to rights-based democratic institutions, processes, and norms. 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.

Paul Nemitz, Principal Adviser on Justice Policy at the EU Commission, addresses the question of what drives democracy. In his view, technology has undoubtedly shaped democracy. However, technology as well as legal rules regarding technology have shaped and have been shaped by democracy. This is why he says it is essential to develop and use technology according to democratic principles. He writes that there are libertarians today who purposefully design technological systems in such a way that challenges democratic control. It is, however, clear that there is enough counterpower and engagement, at least in Europe, to keep democracy functioning, as long as we work together to create rules that are sensible for democracy's future and confirm democracy's supremacy over technology and business interests.

Research associate at the University of Oxford and Professor at European University Cyprus, Paul Timmers, writes about how AI challenges 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.

Tania Sourdin, Dean of Newcastle Law School, Australia, asks: what if judges were replaced by AI? She believes that although AI will increasingly be used to support judges when making decisions in most jurisdictions, there will also be attempts over the next decade to totally replace judges with AI. Increasingly, we are seeing a shift towards Judge AI, and to a certain extent we are seeing shifts towards supporting Judge AI, which raises concerns related to democratic values, structures, and what judicial independence means. The reason for this may be partly due to the systems

used being set up to support a legal interpretation that fails to allow for a nuanced and contextual view of the law.

Pam Dixon, Executive Director of the World Privacy Forum, writes about biometric technologies. She says that biometric technologies encompass many types, or modalities, of biometrics today, such as face recognition, iris recognition, fingerprint recognition, and DNA recognition, both separately and in combination. A growing body of law and regulations seeks to mitigate the risks associated with biometric technologies as they are increasingly understood as a technology of concern based on scientific data.

We invite you to learn more about how our world is changing. As a way to honor this milestone, we have assembled a list of articles from around the world from some of the best experts in their field. This issue would not be possible without the assistance of many people. In addition to the contributing authors, there were many other individuals who contributed greatly. TPQ's team is proud to present you with this edition.

An important acknowledgement goes to our premium corporate sponsor BEKO. In addition, we would like to thank our online sponsor, and the sponsor of this issue, Monaco Economic Board. We would also like to express our appreciation for the continuing support of our other sponsors: Gordon Blair and TEB.

As always, we look forward to your feedback.

Selim Alan
Editor-in-Chief

CEPTETEB İŞTE

@HOME @WORK @MOBILE
WITH YOU, WHEREVER YOU ARE

CEPTETEB
İŞTE

CEPTETEB İŞTE, TEB's business banking app that makes it easy to take care of all of your banking needs wherever and whenever you need to, is now available on mobile platforms too.

TEB

cepteteb.com.tr / 444 0 832

Download on the App Store

GET IT ON Google Play

THE STATE OF AI POLICY: THE DEMOCRATIC VALUES PERSPECTIVE

In a world where data means power, vast amounts of data are collected every day by both private companies and government agencies, which then use this data to fuel complex systems for automated decision-making now broadly described as "Artificial Intelligence." Activities managed with these AI systems range from policing to military, to access to public services and resources such as benefits, education, and employment. The expected benefits from having national talent, capacity, and capabilities to develop and deploy these systems also drive a lot of national governments to prioritize AI and digital policies. A crucial question for policymakers is how to reap the benefits while reducing the negative impacts of these sociotechnical systems on society.

Merve Hickok* & Marc Rotenberg**

TPQ
Winter 2021/22

* Merve Hickok is CAIDP Research Director and Founder of the AIethicist.org.
** Marc Rotenberg is Founder and President of CAIDP and a faculty member at Georgetown Law.

The past year has produced rapid changes in the world of AI policies and practices. National governments and international organizations are moving to create new frameworks to tip the balance in favor of benefits of artificial intelligence. There is consensus that wider adoption of these systems requires a level of trust from public, which in turn requires human-centric systems. But "trustworthy" and "human-centric" cannot simply be slogans. For a system to be trustworthy, it must be subject to independent evaluation. It should produce results that can be verified, replicated, and proven. "Human-centric" necessarily requires the centrality of human agency. Unaccountable, autonomous devices making decisions about people stand at the opposite end of the spectrum from human-centric, trustworthy AI.

From a civil society perspective, so far, the lack of regulations which would ensure oversight over these systems, coupled with minimum transparency have been an increasing source of concern to those working to keep governments, public and private entities accountable. With the expanding adoption of AI systems, the implications over fundamental human rights, democratic values and rule of law are at the core of the discussions. The right to privacy and other rights, including the rights to health, education, due process, freedom of movement, freedom of peaceful assembly and association, and freedom of expression can all be impacted by the way AI systems are deployed.

The Center for AI and Digital Policy (CAIDP) has set forth an objective in 2020 to independently assess global progress toward human-centric and trustworthy AI. Researchers at CAIDP agreed that such progress required an extensive empirical survey of AI policies in countries across the world. The survey also required metrics which corresponded to widely accepted frameworks and principles, allowing for quantification of commitments and practices.[1] The result, AI and Democratic Values Index, has these objectives: (1) to document the AI policies and practices of influential countries, based on publicly available sources, (2) to establish a methodology for the evaluation of AI policies and practices, based on global norms, (3) to assess AI policies and practices based on this methodology and to provide a basis for comparative evaluation, (4) to provide the basis for future evaluations, and (5) to ultimately encourage all countries to make real the promise of AI that is trustworthy, human-centric, and provides broad social benefit to all.

The AI Index is based on such well-established norms and sources of authoritative assessments. The OECD/G20 AI Principles[2] provided a starting point, as did

[1] Marc Rotenberg, "Time to Assess National AI Policies and Practices"*, Communications of the ACM, 24 November 2020,* https://cacm.acm.org/blogs/blog-cacm/248921-time-to-assess-national-ai-policies/fulltext

[2] OECD AI Principles, (2019). https://oecd.ai/en/ai-principles

the Universal Declaration for Human Rights[3], the most widely recognized legal instrument for fundamental rights. We developed a methodology, drawing on the work of international human rights organizations and data protection experts. We revised questions as our work progressed, and new factors were uncovered. We recognized early on the difference between a country's endorsement of a key principle, such as "fairness," and a country's implementation of that principle. Endorsement is easy to measure, implementation, not so much. In highlighting this distinction, we hope others will also look more closely at the difference between what countries say and what they do, all with the larger purpose of closing that gap. And we knew we could not look at the practices of all countries, so we chose those countries (again relying on objective metrics) that we thought would be most impactful.[4] In 2021, CAIDP published its analysis of 30 countries against 12 metrics. In 2022, the analysis has been expanded to 50 countries with the assistance of more than 100 policy experts. Our research team reviewed national AI strategies, commitments to international policy frameworks, adherence to democratic norms such as data protection and transparency, statements from policy makers, and the activities of government agencies. The country report narratives provided the basis for quantitative assessment, followed by ratings and rankings for individual countries.

"How a national government handles the questions of accountability, power and progress is directly relevant to how much it respects democratic values and fundamental rights."

Some AI systems used by public or private actors carry huge risks for society and democratic values if their development and deployment do not center the dignity and rights of natural persons. How a national government handles the questions of accountability, power and progress is directly relevant to how much it respects democratic values and fundamental rights. Of course, no country operates in a vacuum. International organizations, such as the OECD, the G7 and G20, UNESCO, the European Union, and the Council of Europe can establish policy frameworks and legal standards that promote beneficial uses of technologies and curb the dangerous ambitions of some governments. This is particularly evident in the AI policy field where emerging global norms could, if effectively implemented, prevent the use of AI techniques for social scoring, mass surveillance, and autonomous weapons. Therefore, a look at the state of global AI policy also requires an understanding of progress in these organizations.

[3] United Nations, *Universal Declaration of Human Rights.*

[4] CAIDP, *Artificial Intelligence and Democratic Values Index 2020 (2020).* https://www.caidp.org/reports/aidv-2020

From an individual country perspective, this year's edition of AI & Democratic Values Index[5] places Canada, Germany, Italy, and Korea in the top tier for their global leadership on AI policy, their commitment to democratic values, and meaningful engagement with the public on proposed AI strategies. Other important factors for top rankings were a well-established data protection infrastructure, support for algorithmic transparency, and a commitment to fairness, accountability, and transparency for AI systems. In the past year, Canadian authorities determined that ClearviewAI was a form of mass surveillance and violated the privacy and data protection rights of Canadians. Germany continued its leadership on AI policy in the European Union, emphasizing protection for fundamental rights and ongoing public participation on AI policy development. As host of the G20 summit, Italy advanced AI policy proposals, emphasizing data protection and gender equality, diversity, and inclusion. Korea introduced new requirements for AI impact assessments, published guidance on Personal Information Protection, and expanded algorithmic transparency.

China has also adopted sweeping new laws for both data protection and the regulation of recommendation algorithms. Although the privacy rules look very similar to the GDPR and the regulation for the governance of recommendation algorithms share similar ambitions to those proposals pending in both the European Union and the U.S Congress, there are real concerns about AI policies that are intended to favor the government in power. In that context, the goals of transparency and accountability is offset by the inherent bias of such a legal structure.

Despite its private sector's global edge in AI innovation, the U.S. ranked in third tier in the Index. The absence of a legal framework to implement AI safeguards and a federal agency to safeguard privacy raises concerns about the ability of the U.S. to monitor AI practices.

On the global front, the UNESCO Recommendation on the Ethics of AI, adopted by 193 countries in November 2021, was the single most significant AI policy development of the past year.[6] The UNESCO Recommendation speaks directly to the widespread – and widely shared – aspiration of countries that AI should benefit humanity. In the AI policy field that barely existed a few years ago, the UNESCO AI Recommendation is a remarkably comprehensive AI policy framework, touching upon established concerns regarding AI systems, such as fairness, accuracy, and transparency, and emerging issues, including gender equity and sustainable development. UNESCO's proposal for Ethical Impact Assessment provides a

[5] CAIDP, *Artificial Intelligence and Democratic Values Index 2021 (2021).* https://www.caidp.org/reports/aidv-2021

[6] UNESCO, *Recommendation on the ethics of artificial intelligence* (2021), https://en.unesco.org/artificial-intelligence/ethics

powerful new tool to assess, in advance, the consequences of the deployment of AI systems. Recognizing the importance of the first global framework for AI ethics, CAIDP has this year altered one of the metrics to take account of the significance of the UNESCO Recommendation on AI. It is a development worth acknowledging and celebrating. In future reports, we will likely add another metric to assess the far more challenging issue of implementation.

"As the field of AI policy rapidly matures, we observe the growing presence of judicial decisions, now shaping the laws of algorithms. In several cases, including the secretive evaluation of employee performance, courts have rejected opaque automated decisions. These judgements are based on well-established legal frameworks, such as the GDPR, though we see also legislative efforts to make automated decision-making with AI techniques more accountable."

Since the publication of our last report, we also note the introduction of the European Commission proposal for the regulation AI. The Commission has set out a comprehensive, risk-based approach that could extend the "Brussels Effect" to the global governance of AI. The European Parliament has also signaled its intention to strengthen key provisions, and likely will prohibit the use of AI techniques for remote biometric identification. Meanwhile, the Council of the European Union, under the Presidency of Slovenia and now France, have proposed additional texts that would among other changes, extend the prohibition on social scoring to private companies as well as public agencies.

2021 also marked the adoption of Resolution 473 in Africa, concerning the need to undertake a study on human and peoples' rights and artificial intelligence. The African Commission on Human and Peoples' Rights called on State Parties "to ensure that the development and use of AI, robotics and other new and emerging technologies is compatible with the rights and duties in the African Charter and other regional and international human rights instruments, in order to uphold human dignity, privacy, equality, non-discrimination, inclusion, diversity, safety, fairness, transparency, accountability and economic development as underlying principles

that guide the development and use of AI, robotics and other new and emerging technologies."[7]

The Council of Europe (COE), continent's leading human rights organization[8], comprised of 47 member states, had established the Ad Hoc Committee on Artificial Intelligence (CAHAI) in September 2019.[9] The mandate of the CAHAI was to "examine the feasibility and potential elements on the basis of broad multi-stakeholder consultations, of a legal framework for the development, design and application of artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law."[10] The CAHAI held its final meeting in December 2021.[11] At the end of the meeting, the CAHAI adopted the "Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law." The CAHAI framework contains an outline of the legal and other elements which in the view of the Committee could be included in legally binding or non-legally binding instruments of COE that will make up an appropriate legal framework (possibly a convention) on AI. The CAHAI framework is now submitted to the Committee of Ministers for further consideration. The Committee has already emphasized in 2021 that such systems should be developed and implemented in accordance with the principles of legal certainty, legality, data quality, non-discrimination, and transparency.[12]

We also noted the growing conflict over the deployment of facial recognition for mass surveillance. While the European Parliament voted to ban the use of AI technology for this purpose, many governments and private companies pushed forward new systems for surveillance in residential communities, inside school classrooms, and at public parks. These are not the CCTV cameras of old, but sophisticated image processing systems, designed specifically to identify individuals in public spaces by name. In some countries, this system of unique identification is then tied to elaborate government databases for scoring people based on their allegiance to

[7] African Commission on Human and Peoples' Rights, *473 Resolution on the need to undertake a Study on human and peoples' rights and artificial intelligence (AI), robotics and other new and emerging technologies in Africa - ACHPR/ Res. 473 (EXT.OS/ XXXI)* (25 February 2021).

[8] Council of Europe, "Who we are," https://www.coe.int/en/web/about-us/who-we-are

[9] Council of Europe, "The Council of Europe established an Ad Hoc Committee on Artificial Intelligence – CAHAI," (11 September 2019), https://www.coe.int/en/web/artificial-intelligence/-/the-council-of-europe-established-an-ad-hoc-committee-on-artificial-intelligence-cahai

[10] Council of Europe, "CAHAI - Ad hoc Committee on Artificial Intelligence," https://www.coe.int/en/web/artificial-intelligence/cahai

[11] Council of Europe, "The CAHAI held its 6th and final plenary meeting," (2 December 2021), https://www.coe.int/en/web/artificial-intelligence/-/outcome-of-cahai-s-6th-plenary-meeting

[12] Council of Europe, "Declaration by the Committee of Ministers: the use of computer-assisted or AI-enabled decision making by public authorities in the area of social services must respect human rights" (17 March 2021), https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset_publisher/csARLoSVrbAH/content/declaration-by-the-committee-of-ministers-the-use-of-computer-assisted-or-ai-enabled-decision-making-by-public-authorities-in-the-area-of-social-servi

the government in power. It is a form of social control beyond the imagination of even Jeremy Bentham's panoptic prison design.

We called attention also to the unfortunate failure of negotiators at the UN conference in late December to make progress on a proposal to limit, or better to prohibit, the use of lethal autonomous weapons. This occurred in the same year that the United Nations suggested the first use of autonomous drone swarms to target and kill retreating military forces in the civil war in Libya.

As the field of AI policy rapidly matures, we observe the growing presence of judicial decisions, now shaping the laws of algorithms. In several cases, including the secretive evaluation of employee performance, courts have rejected opaque automated decisions. These judgements are based on well-established legal frameworks, such as the GDPR, though we see also legislative efforts to make automated decision-making with AI techniques more accountable. We report these outcomes favorably as algorithmic transparency remains one of our key metrics for the evaluation of AI policies and practices.

In addressing the need to advance democratic values in the age of AI, the ability of the European Union, the United States, and allies to work in common purpose remains central. On that front, the past year provides reason for both optimism and concern. The EU and the U.S. launched a Trade and Technology Council in 2021 that set out a common framework on AI policy that could promote further transatlantic cooperation.[13] The good news is that "human rights" and "democratic values" undergird many of the proposals. Top officials in the Biden Administration also expressed support for the EU AI Act, a key legislative framework that will likely move forward in 2022.

At the same time, the future of the EU AI Act is not certain, as some politicians have made the mistake of assuming it is possible to trade the protection of rights for innovation. Technologies that fail to protect rights are not innovative, they are oppressive and stifling. On the U.S. side, several federal agencies have initiated AI-related "listening sessions," but the necessary work of establishing legal standards to protect democratic values has yet to begin.

Still, our survey of national AI policies and practices also revealed the hard work of many NGOs, advocates, academics, and government officials, around the world, who have fully engaged the challenges that AI poses and are prepared to stand on

[13] European Commission, *EU-US Trade and Technology Council: Commission launches consultation platform for stakeholder's involvement to shape transatlantic cooperation* (October 2021). https://ec.europa.eu/commission/presscorner/detail/en/IP_21_5308

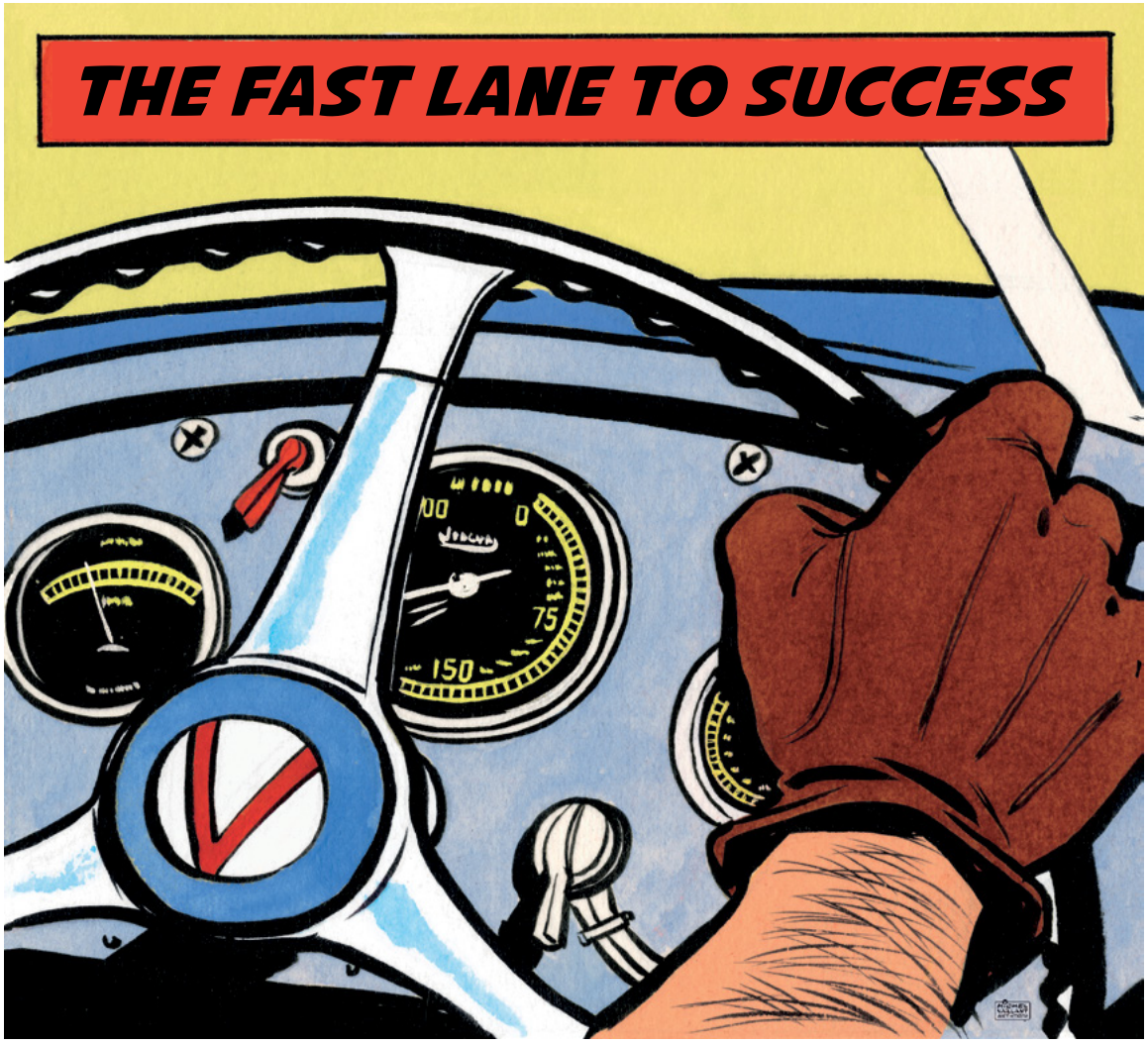
the front lines in defense of fundamental rights. The remarkable progress made by the "ReclaimYourFace" campaign in Europe, and similar campaigns in Africa, Asia, and Latin America speak to a rapidly growing public recognition that not all technologically transformative impacts should be welcome.

This point was made clear during the past year with the call from Michelle Bachelet, United Nations High Commissioner on Human Rights, for a prohibition on AI techniques that fail to comply with international human rights law. Commissioner Bachelet stated, "The higher the risk for human rights, the stricter the legal requirements for the use of AI technology should be."[14]

There is a growing understanding that "red lines" are necessary to safeguard fundamental rights. And in that recognition may be found also the key to aligning AI policies and practices, to narrowing the gap between the world of AI as it is and the world of AI we wish to inhabit. If AI is to remain human-centric, then we must determine the appropriate applications of AI.

[14] United Nations, *Urgent action needed over artificial intelligence risks to human rights* (September 2021). https://news.un.org/en/story/2021/09/1099972

GORDON S. BLAIR
LAW OFFICES



MASTERED LEGAL AND **TAX SOLUTIONS**

MONACO • GENEVA

www.gordonblair.com
7 rue du Gabian, BP 449 - MC 98011 Monaco cedex - T +377 93 25 85 25
Route de Pré-Bois, 29 CP 231 - 1215 Genève 15 - T +41 22 799 45 00

ETHICS OF AI AND DEMOCRASY: 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*

TPQ
Winter 2021/22

* 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 Welser, *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 Poullet, "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 opaqueness 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 Akın 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

ULUSLARARASI İLİŞKİLER
Akademik Dergi

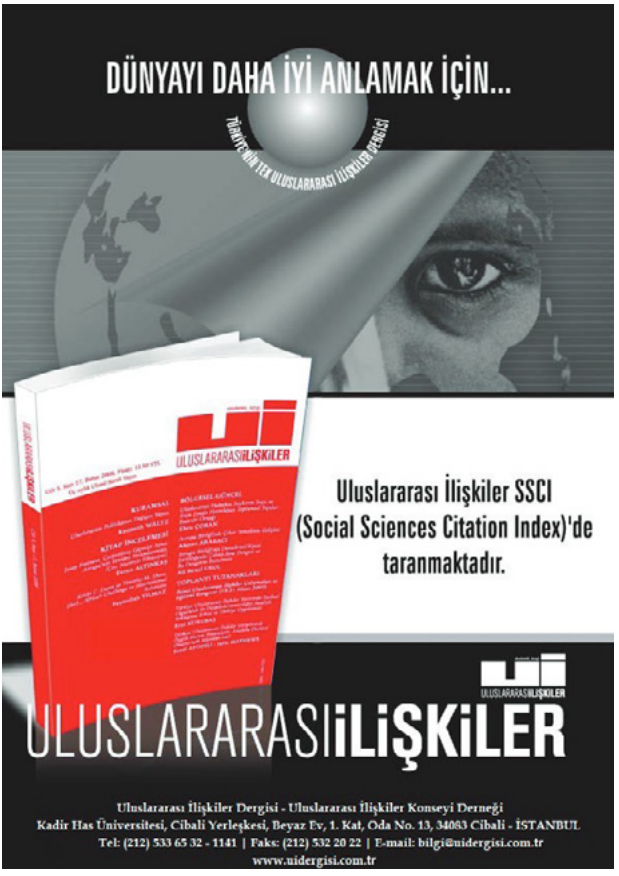
CALL FOR PAPERS

ULUSLARARASI İLİŞKİLER
(INTERNATIONAL RELATIONS)

Published in association with the *International Relations Council of Turkey*

The Editorial Board of **Uluslararası İlişkiler** *(International Relations)* Journal invites submissions of papers for its forthcoming issues. The journal publishes articles on diplomatic history, international relations theories, international law, political economy, regional/current issues and strategic issues. The Journal is covered by Social Sciences Citation Index (SSCI). All articles submitted to the journal are confidentially refereed by at least two independent referees.

Further information is available at www.uidergisi.com.tr



PEOPLE OR TECHNOLOGY: WHAT DRIVES DEMOCRACY?

There is no doubt that technology has shaped democracy. However, it is also the legal rules concerning technology which have shaped and have been shaped by democracy. Developing and using technologies according to democratic rules is what we need. There are libertarians today who deliberately design technological systems in a way that makes control by democratic states increasingly difficult. However, there is sufficient counter power and engagement, at least in Europe, to ensure that democracy functions and we work together to produce rules that are sensible for the future of democracy and confirm the supremacy of democracy over technology and business interests.

Paul Nemitz*

TPQ
Winter 2021/22

* Paul Nemitz is the Principal Adviser on Justice Policy in the EU Commission; Member of German Data Ethics Commission, Global Council on Extended Intelligence; Visiting Professor of Law, College of Europe. He here expresses a personal opinion, not necessarily that of the European Commission.

In a democracy, the path to change must be open. This involves changing both the government and the people at the top. Change is brought about by people and thus history is not predetermined by ideologies or technological advancements. The future is open, some harm will be unavoidable, but nothing good can happen unless people engage with rigor and perseverance. We must make this clear because both authoritarians and technocrats claim the opposite. In the perception of those who believe the future is predetermined, what they foresee and plan will unavoidably happen. That is their strategy for disengaging people from democratic involvement. However, we shouldn't let this happen.

There is no doubt that technology has shaped democracy. However, it is also the legal rules concerning technology which have shaped and have been shaped by democracy. For example, Hitler used public radio to manipulate public opinion. It was for this reason that many countries in Europe passed legislation after World War II establishing detailed rules for public radio and public television, including their administration and financing. The objective of these rules was and is to ensure that these technologies serve democracy in the form of public broadcasters, independent of profit or state control. The rise of populism and antidemocratic tendencies can be linked to a lack of strong, independent public broadcasters and rules that ensure the media remain pluralistic and work independently in order to inform the public in a way that fosters democracy. Democracies have always intended to use media policy to make sure that a genuine democratic public sphere (Habermas) is created and that it remains.

In the United States, the degeneration of democracy is traced back to private television becoming so powerful and so vital for campaigning that candidates for elected office are either required to be very rich or to spend an increasing part of their time collecting funds to finance TV advertisements. Former Vice President Al Gore once told me that when elected representatives had to consider the impact each call would have on campaign financing, politics was no fun anymore. The Harvard Constitutional Law Professor Lawrence Lessig has described in detail the poison money has brought into American politics,[1] and much of this need for money has to do with the development of Television and Radio financed through advertisements, for which candidates for office must pay to be able to win an election campaign.

We have seen a surprising pattern of political advertising spending on Facebook during the U.S. election campaigns, and now we have arrived at contemporary AI-driven social networks and their effects on democracy. Former President Trump

[1] Lawrence Lessing, "Lesterland" TED Talk on Youtube: https://www.youtube.com/watch?v=mw2z9lV3W1g

had to pay a lot less per view on Facebook than democratic candidates. How come? Because the Democratic Party initially had many candidates, they all vied for the same electorate during the primaries. As a result of collecting mass data and leveraging artificial intelligence, Facebook auctions off every American voter with a profile that is of interest to the bidder bidding the most. And there were many bidders on the democratic side, namely all the Democratic Party presidential candidates. This increased the auction price and thus the spending of democratic candidates. Trump, on the other hand, as the sole candidate of the Republicans, benefitted from a lower price as nobody in the add auctions stood against him when he was reaching out to his potential voters.[2]

"All of these examples are American, and for a very good reason: the lack of rules on the collection and use of personal data in the U.S., as well as an under regulated technology market and an advertisement market on fire, has a terrible impact on democracy."

Even worse, individual micro-targeting of social media messages to voters, calibrated for maximum manipulation with the help of mass data and artificial intelligence, robs the election process of the transparency and scrutiny that democratic societies need. When only the addressee sees the messages received from the party, the party can say whatever it wants without being held accountable for the inconsistency of its messaging. Because democracy cannot work in this way, the European Commission has proposed a regulation of the election advertising on the internet.[3] Could we leave this matter to the market or civil society to resolve? Dream on. The state must take responsibility for the institutional guarantees of democracy to function, including the institutional guarantees for the public sphere. It is good that the American Journalist Julie Angwin and her philanthropically financed tech accountability newsroom "The Markup" developed a "Citizenbrowser" app plugin for volunteers to document the divergence of messaging based on micro targeting on Facebook.[4] Her findings confirm the need for democracy to act through law.

In the U.S., it may be fair to say that every new technology cycle relevant to media and public opinion shaping has driven up the price of democracy in that evermore only the richest and smartest have the time and money to in fact find the truth and

[2] Jeremy B. Merill, "Facebook Charged Biden a Higher Price than Trump for Campaign Ads," *The Markup*, 29 October 2020, https://themarkup.org/election-2020/2020/10/29/facebook-political-ad-targeting-algorithm-prices-trump-biden

3 "European Democracy: Commission sets out new laws on Political Advertising, Electoral Rights and Party Funding," *European Commission*, (25 November 2021), https://ec.europa.eu/commission/presscorner/detail/en/ip_21_6118

[4] "Citizen Browser," *The Markup*, https://themarkup.org/series/citizen-browser

make proper judgments among the cacophony of fake news and lies and only rich people or those who made the necessary pleasing gestures to bring in money have a chance to be elected.[5] Since the first Obama campaign, billions of dollars have been spent on developing the app and algorithmic environment for voter mobilization. On the day of the election in 2012, the mobilization app of the Mitt Romney campaign did not work properly.[6]As early as midday, Obama staffers at every polling station knew who they still needed to call or pick up to come to the election because the system they had in place to track voters at the polling stations against their list of potential Obama voters worked. They could thus mobilize people until the very last minute. Mitt Romney could not, because his app failed.

As well as the good arguments, this campaign may have been the first to win through technology. The following campaigns saw technology become an increasingly important element, culminating with the Cambridge Analytica Scandal of the Trump Campaign, which showed how Facebook made it possible for the Trump Campaign to use personal information and manipulate people. All of these examples are American, and for a very good reason: the lack of rules on the collection and use of personal data in the US, as well as an under-regulated technology market and an advertisement market on fire, has a terrible impact on democracy. Currently, elections and democracy in America are large money feasts for big businesses, which perverts democracy. There is a very real chance that in America's campaigns of the future, the best arguments will matter less, and the one with the most sophisticated AI and Tech systems will win. This means that the campaign with the most money will have a greater chance of winning, since more money probably buys better manipulation technology.

What keeps campaigns in Europe away from these perversions are two sets of rules: Election and party financing are often provided by the state, at least to some extent. This and stringent rules limiting the financial influence of big capitalists as well as restrictions on data collection and use are key. An effective democracy cannot function without well-calibrated and rigorously enforced rules. The relationship between Tech, AI, and Democracy is heavily influenced by them, AI is being the key tool for collecting and manipulating individual data.

Can AI and Blockchain Drive down the Cost of Democracy and Re-democratize Elections?

In an optimistic vision, AI in the future will help curate reliable, true information,

[5] Lawrence Lessig, "Lesterland" video: https://www.youtube.com/watch?v=mw2z9lV3W1g

[6] Steve Mullis, "On Election Day, Romney's Killer Whale 'App' Couldn't Stay Afloat," *NPR*, 10 November 2012, https://www.npr.org/sections/alltechconsidered/2012/11/10/164869691/on-election-day-romneys-killer-whale-couldnt-stay-afloat?t=1642517005525

while the Blockchain will ensure efficient, secure data inputs and transmissions. Additionally, since Blockchain relies on decentralized consensus to pass along information, it contains an element of democracy. We see many discussions about technologies without a solid understanding of the power consolidation and economic profit objectives that motivate interest in developing and deploying technology like AI and Blockchain. It is clear that much of the conceptual discussion on AI Ethics or the AI and Democracy debate or even the "AI for good" or "Blockchain for good" debate suffer from this malaise. In this sense, techno naivety makes it easier for big tech to push "techno solutionism" in which Tech for good is held up as solving all problems, including challenges to democracy. Technology is certainly capable of solving many problems, hopefully including those that it created in the first place. It is also certain, however, that if Big Tech is in control of this endeavour, or even if it runs it, it will not lead to results those democratic legislators or the majority of people would consider good. Because capitalists cannot escape the stock market's pressures. In other words, those pressures are not for the best, but rather for the profit of the shareholders and for monopoly power.

> *"AI will cement conservative structures of society and make reforms, based on criticism and democratic will, much more difficult. We must be careful not to lock cities, governments, and other public services into AI systems that have huge sunk costs that will make any subsequent reversal of policy difficult."*

Developing and using technologies according to democratic rules is what we need. The definition of Tech for Good is subject to many different points of view, and some tech may be good for some and not for others. It is a classic case of democratic deliberation and decision making when it comes to "Tech for Good." It appears in principle that Big Tech today accepts democracy as the supreme authority over technology, at least in op-eds and college commencement speeches. Nevertheless, profits in capitalism are used to generate more profits, and that means lobbying democracy to death in order to make sure that no rules by legislators reduce profits. Is big tech always successful in this endeavor? Praise God it is not, thanks to the engagement of many. But it is a constant battle. Furthermore, we must ensure that all the structures which keep democracy alive and curtail the impact of big money and technology on democracy are strengthened simultaneously with the exponential financial growth and dominance of big tech in public discourse. It seems that the more money GAFAM (Google, Apple, Facebook, Amazon and Microsoft) spends to

buy academics, journalists, or entire publishing houses, politicians, and competitors, the more important it becomes to strengthen our institutions and rules to make sure this does not happen and that the checks and balances that democracy requires are maintained and strengthened.

Undermining democracy's system of checks and balances is the single most significant risk AI and its corporate backers pose to democracy. The claim that AI will be a better curator of truth than journalists and a better writer of language undermines the critical potential of the press as the fourth estate in a democratic society. By replacing journalists with automated algorithms, which is cheaper, the press will lose its critical, questioning function against power, whether public or private. Past facts are the basis of artificial intelligence. They do not tell us anything about how the world should be, these facts cannot provide a basis for normative views and criticisms. Consider this in conjunction with the erosion of revenue from advertising in the press, to the benefit of Google and Facebook. Only these two companies today reap 80 percent of the new internet advertisement revenues previously accrued to hundreds of newspapers. In turn, they hand-feed the press in their so-called news programs, making them dependent on them, and rendering them friends rather than critics of their power. Eventually, they will take them over or imitate them. Amazon founder Bezos took over the Washington Post. Next up I bet: Springer in Germany and the United States, with KKR as one of its major shareholders. What will they do to maximize return on investment if they don't sell to one of the GAFAMs?

Asserting that artificial intelligence (AI) will be better scientists than human scientists and that we should put our trust in "trustworthy AI" is insulting to human thinking, reflection, and critical thinking which has been developed through enlightenment. Using data from yesterday, AI will cement conservative structures of society and make reforms, based on criticism and democratic will, much more difficult. We must be careful not to lock cities, governments, and other public services into AI systems that have huge sunk costs that will make any subsequent reversal of policy difficult. Though public officials are known for their persistence and resistance to change, they still learn quicker and are better at following new directions after an election which led to a change in government than machines. If a child has seen the door handle three times, he or she understands that it opens by pushing down on it. AI needs 100000 thousand sets of training data for even the simplest tasks. And even if that distance between human learning and artificial intelligence becomes smaller, we cannot have public administrations that are devoid of human faces, empathy, and good judgement, as well as deliberation, reasonable discretion, and all that is genuinely human. In the end, humans cannot be controlled by machines, and it is no paradise to have AI take ever more decisions over our lives. If

we become objects of machine governance and control, human free governance will be hell for humans, since it will inevitably reduce their humanity. The scarcest and most needed resource today is not the new technology, but the willingness and ability to agree on the rules according to which we want to live together. Covid is the latest example for that. And we need to work on this challenge, within societies and across the Atlantic.[7]

The technological society in which we already reside and which we are becoming ever more requires the participation of many in democratic processes. It involves a willingness to fight for freedom, fundamental rights, the rule of law, and democracy in the age of artificial intelligence. To achieve this, we must let go of naive hopes for an AI which we can trust, corporations that are benevolent, and the idea that we don't need to rely on democracy anymore, which sets rules for technology and enforces them. The Blockchain and crypto scene is going very far in this direction. There are libertarians today who deliberately design technological systems in a way that makes control by democratic states increasingly difficult. Those who follow these trends do so either because they are fascinated with the technology or to make money. Others even pursue an idealistic agenda, in that they hope these technologies will solve major issues of sustainability. Despite this, there are those, I fear they are among the economically most successful shapers of the crypto-sphere and Blockchain, who believe they can continue to make the biggest profits if the systems are designed to systematically evade laws and control by democratic governments. They may follow a public discourse of respect for legislation and democracy, like big tech. However, their actions speak for themselves. Again, let us look at the economics. Within only a few years, the major crypto exchanges have grown to a capitalization similar to those of the world's largest banks. As of today, the question is no longer whether the banks will buy crypto exchanges, but whether the crypto exchanges will buy the banks. Together, AI and crypto will become a very powerful combination of technology and money. Will it be beneficial for democracy? I have serious doubts. We cannot leave the development and deployment of these technologies without the necessary rules and controls imposed by democracy.

The primacy of democracy is increasingly challenged by technology. Blockchain and AI will be the strongest challenge since they combine powerful technology with a lot of money. In contrast to AI, Blockchain also attracts millions of people who believe they can earn money with these new cryptocurrencies. The number of small investors in Cryptocurrency, as well as the students in the global teaching courses, is growing by the millions, just as the profits of the crypto exchanges and promoters

[7] Paul Nemitz, "Democracy Through law: The Transatlantic Reflection Group and its manifesto in defence of democracy and the rule of law in the age of "artificial intelligence," *European Law Journal*, (1 December 2021): p. 1-12, https://doi.org/10.1111/eulj.12407

of cryptocurrency are growing by the billions. Democracy has nothing to do with consensus on Blockchain, and decentralization in cryptography is not what defines democracy. It lacks transparency, outside criticism, renewal, and inclusive discussion with all, not just those who invest in computers for mining. A blockchain is a democracy in which only the rich who invested in mining can vote, and the richer get more votes because computing power determines whether a consensus is reached. Not to mention the power of the promoters and exchanges and their profits. Blockchain cannot be used to build a society, but, like AI, it can fulfil certain defined functions that are regulated by law.

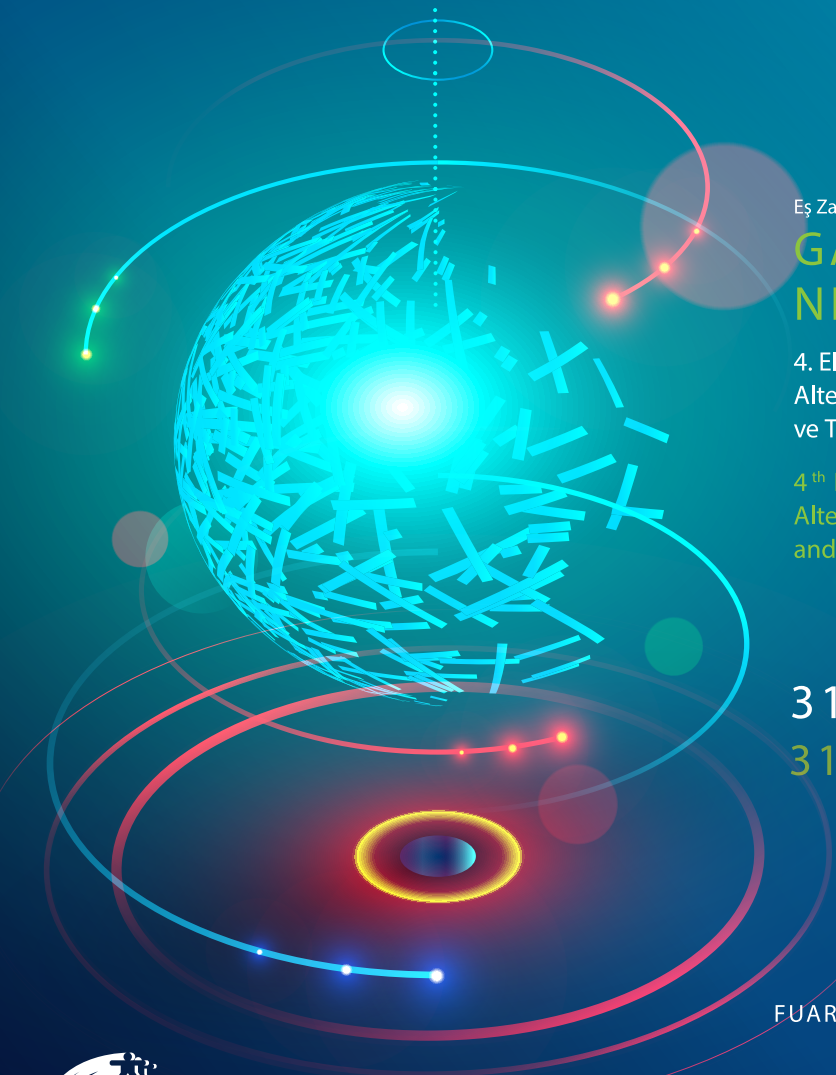
The way must be open for change that is what democracy is about. For decades, neo liberalism and conservatives have failed to regulate the internet and Blockchain. These times are now over. The change now is that the European Commission has put on the table a package of proposals for binding rules for the new technologies and networks,[8] with the aim to secure Democracy, Rule of Law and Fundamental Rights for the future and to make the new markets work with openness for future entry and innovation. Of course, these rules shall also serve innovation and economic growth. Of course the lobby to soft wash these rules so that they don't serve their purpose and only serve one goal, namely private profits, is on, while the legislators deliberate. However, there is sufficient counter-power and engagement, at least in Europe, to ensure that democracy functions and we work together to produce rules that are sensible for the future of democracy and confirm the supremacy of democracy over technology and business interests. We must recognize that law is the most noble expression of democracy, and stop talking it down. It is necessary to foster human engagement in deliberation based on language and a critical attitude towards power, rather than preaching technocratic visions of the world, in which democracy plays a smaller role, and which could potentially attract both technocrats and authoritarians. The future is made by humans, so the way must remain open to change. Are AI and other technologies going to support this transition to a more humane, more democratic society? No one knows for sure. Tax-financed research programs are on their way on both sides of the Atlantic for this purpose. Tech Giants spent little for this purpose and continue to encourage views that are hostile toward the law and strong democracy, which makes binding decisions and enforces them. They lobby one way on specific legislation, while in the other was publishing books and articles vowing allegiance to democracy. However, we should not wait for technology or corporations to repair democracy. Let's make democracy work again by engaging people.

[8] See for example the proposed rules on AI at https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence and the proposed DSA at https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

PETROLEUM ISTANBUL

15. Uluslararası Petrol, LPG, Madeni Yağ, Ekipman, İstasyon Market Ürünleri ve Teknoloji Fuarı

15th International Petroleum, LPG, Lubricants, Equipment, Convenience Store Products and Technologies Exhibition



Eş Zamanlı Fuar Concurrent Exhibition

GAS&POWER NETWORK

4. Elektrik, Doğal Gaz ve Alternatif Enerji, Ekipmanları ve Teknoloji Fuarı

4th Electricity, Natural Gas and Alternative Energy Equipment and Technologies Exhibition

31 Mart - 2 Nisan
31 March - 2 April
2022

TÜYAP
FUAR VE KONGRE MERKEZİ
TÜYAP EXHIBITION and CONGRESS CENTER

petroleumistanbul.com.tr
gaspowernetwork.com
efo.com.tr

E·N·E·R·J·İ FUARCILIK
www.efo.com.tr

KOSGEB

Bu fuar 5174 sayılı kanun gereğince TOBB (Türkiye Odalar Borsalar Birliği) izni ile düzenlenmektedir.

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

WHAT IF JUDGES WERE REPLACED BY AI?

Although it is likely that in most jurisdictions AI will be increasingly used to support judges when making decisions, it is also probable that there will be moves over the next decade to entirely replace judges in some areas with forms of AI. This shift towards Judge AI, and to some extent even shifts towards supportive Judge AI, raise issues that impact democratic values, structures and may reshape any notion of judicial independence. This is partly because the systems that are used may be designed to support an interpretation of the law that will not permit a nuanced and contextual approach. Instead, the systems may reflect political or other agendas with little or no independent oversight.

Tania Sourdin*

TPQ
Winter 2021/22

* Professor Tania Sourdin is the Dean of Newcastle Law School, Australia. Parts of this article are drawn from Tania Sourdin, *Judges, Technology and Artificial Intelligence* (UK: Edward Elgar, 2021).

Whilst some may argue that Judge AI may be more predictable and accurate than human judging, there are numerous questions relating to transparency in decision making, procedural justice and issues relating to substantive justice when data and systems are used to determine a dispute particularly if they are developed outside the judicial system. At present, as digitization of court records and general dispute records is at an early stage in most jurisdictions, there are short term risks that Judge AI will result in perverse outcomes, as the only data to be considered is from cases finaliszed by judges which form only a small percentage of disputes that are finalized. From a longer term perspective, the risk is that faster and cheaper outcomes may be obtained at the expense of justice and the ongoing development of the law.

One significant issue emerges in societies where independent judges assist to develop the law. They do so in many democratic countries by dissenting, developing, innovating and considering law and justice within changing contexts. Whilst it is likely that AI can assist with this task (supportive judge AI), it is not likely, at least in the shorter term, that Judge AI (which involves the complete substitution of AI) will result in creative or innovative outcomes that can aid in the development of the law. Also, although justice is difficult to define, the growth in therapeutic jurisprudence, problem solving courts as well as the recognition that many human disputes cannot be defined by reference to legal issues alone means that AI does not and cannot currently address a myriad of human justice issues. This reality also means that a human being, who does much more than sort, categorize and produce a written outcome to support justice in both criminal and civil dispute arenas is unlikely to become redundant. The human face of justice, as characterized by a judge, includes empathy, compassion, reflection and the acknowledgement and support of the human struggle.

Democracy and Judge AI

Whilst objections to Judge AI can be summarized as including: (i) a missing 'human element'; (ii) transparency issues; (iii) issues associated with system transformation and change; (iv) the challenges of coding law; (v) issues surrounding Judge AI in innovative and novel situations; and (vi) issues associated with creating new law in precedent based systems, perhaps the greatest concerns where judges are replaced by AI relate to threats to democratic systems of government.[1]

In most democratic countries, the concept of separation of powers (sometimes

[1] Paul Nemitz, "Constitutional democracy and technology in the age of artificial intelligence," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 376, No. 2133 (2018): p. 10-13. http://doi.org/10.1098/rsta.2018.0089

called the Westminster system in the UK) is intended to ensure that no branch of government becomes too powerful. In such systems, the judiciary is one of the three arms of government comprised of the judiciary, the executive and the legislature. The judicial arm is intended to support a system of checks and balances, to enforce legislative requirements and to also support the legality and appropriateness of government decision making. Whilst some might question the extent to which the doctrine is realistic,[2] constitutional arrangements can require that each of the three arms of government have distinct roles and function independently of the other.

"In countries where Judge AI is developing quickly, such as China, this may mean that it is unlikely that the position put forward by the State or a powerful party can be challenged, and inequity in terms of resources or coding variables may be more extensive."

It is in this context that additional concerns regarding Judge AI may arise. Most commentators suggest that central to the establishment and maintenance of the rule of law is the concept that the judiciary remains independent and the extent to which the judicial arm is 'independent' may be questionable, particularly if other arms of government are involved in the creation of Judge AI. Supportive Judge AI arrangements could also be perceived to be problematic in some countries, particularly if such supports are regarded as having an 'improper' influence on a judge.

A central tenet of judicial independence that is also challenged by Judge AI relates to how an AI Judge could really be said to serve as a significant 'check' on the executive and legislative branches of government.[3] For example, Zalnieriute and Bell have identified the potential for automation to undermine the independence of the judiciary from interference or usurpation by the other branches of government.[4] In this regard, there are a number of assumptions that are made about automation and these include the view that the judicial role is only to 'robotically apply the law to the facts.'[5] Judges do far more than this, and an analysis of the judicial role and function indicates their important role in settling disputes as well as determining

[2] See discussion in Mike McConville and Luke Marsh, *The Myth of Judicial Independence* (Oxford University Press, 2020).

[3] Andrew C Michaels, "Artificial Intelligence, Legal Change, and Separation of Powers," *Cincinnati Law Review*, Vol. 88, No. 4 (2020): p. 1084.

[4] Monika Zalnieriute and Felicity Bell, "Technology and the Judicial Role," in *The Judge, the Judiciary and the Court: Individual, Collegial and Institutional Judicial Dynamics in Australia* ed. Gabrielle Appleby and Andrew Lynch (Cambridge University Press, 2021).

[5] Michaels, (2020), p. 1084.

them.

There are other issues linked to judicial independence that are not related to potential executive control or interference. For example, there are concerns regarding digital overreach in that many digital corporations that will develop Judge AI may not understand neither democratic notions, nor the rule of law, and may reshape judicial activities in unforeseen ways. The law as code movement is gathering momentum around the world and there are questions about in whose interests a mechanical application of the law might serve:

> 'To be very clear: requiring that law be either as precise as code or be rewritten as fast as code is updated is simply anti-democratic, as this ignores the need for deliberation and compromise in democracy as well as the time required for due process under the rule of law.'[6]

Online Courts and Judge AI

To some extent, at least at present, Judge AI relies on there being written rather than oral evidence which is less likely to be the subject of exploration and interrogation. In countries where Judge AI is developing quickly, such as China, this may mean that it is unlikely that the position put forward by the State or a powerful party can be challenged, and inequity in terms of resources or coding variables may be more extensive. In jurisdictions where there are significant social issues that may have political ramifications, an independent judicial inquiry can be an important mechanism to support a functioning democracy.

In addition, online 'courts' that are created as separate tribunals or entities may raise issues of independence, particularly where they are developed and maintained without judicial supervision or monitoring. For example, both online courts and Judge AI raise concerns about the extent to which judicial independence is supported where court processes are not 'open'. In this regard, judicial independence, impartiality and 'open' courts are often perceived to be essential to the maintenance of the rule of law. Further, some commentators consider those online court developments which involve ODR constitute a potential encroachment on judicial independence by limiting the matters that may come before a judge or by otherwise impacting on the business of courts and therefore judges.

> "When justice slips out of sight … .the prospect of arbitrary, incompetent or unlawful conduct raises its head. Again, if we simply accept the argument that private online dispute resolution is the way in which the majority of disputes,

[6] Nemitz, (2018) p. 9.

> and in some areas all disputes, may be resolved in future we accept this loss of accountability; we further accept the growth of a democratic deficit. And the same is the case if we divert public justice to an unobservable online forum. Our digital courts must be open courts."[7]

Some of these challenges can be overcome by a clear and consistent focus on using technology to support 'open' court hearings however the use of online hearings and hearings 'on the papers' raise issues about the extent to which procedural and participatory justice is supported.

"It seems clear that, in the coming years, technological developments will impact on the role of judges within democratic and non-democratic societies in terms of the retention of human judges, the function of judges, the definition of a court, the 'openness' of justice and also the humanity of the justice system."

Conclusion

Is it realistic that AI will replace anything other than simplistic judicial decision making? My response is yes, although it may take some time and there are many hurdles to overcome and also questions about why we would want to do so.[8] Inevitably however, there is room for improvement. Inconsistent, slow, laborious justice processes can be unjust and are inappropriate in a modern world where AI offers options to enable justice improvements to thrive.

It seems clear that, in the coming years, technological developments will have an impact on the role of judges within democratic and non-democratic societies in terms of the retention of human judges, the function of judges, the definition of a court, the 'openness' of justice and also the humanity of the justice system. However, these impacts need not be negative. In many ways, technology can enable a better appreciation and understanding of the judicial role and judges can play a central role in the redesign of the justice system to ensure that technological reform efforts reflect current and potential future judicial functions. Ensuring that judges are engaged in

[7] Justice Committee, *Court and Tribunal Reforms* (House of Commons Paper No 190, Session 2019), p. 50 [155].

[8] There are many reasons to retain judges that that include maintaining democratic governance arrangements. However, the author notes that a number of scholars consider that this could theoretically mean that 'old' judges might not need to retire (although this of itself raises other ethical and moral questions). See also Yana B. Feygin, Kelly Morris and Roman V. Yampolskiy, "Uploading Brain into Computer: Whom to Upload First?" *Computer Research Repository* (November 2018), CoRR abs/1811.03009.

reform is necessary, not only because of the issues that can emerge in democratic countries that relate to the separation of powers doctrine, but also because judges on the whole are more likely to be able to manage the balancing process that must be undertaken to ensure that the key tenets of the rule of law and inherent in the formulation of justice are maintained and supported.

This means that the design of the justice system of the future must incorporate a nuanced understanding of the role of judges in society and enable judicial activism in an extended public form to assist to guide that design process. In addition, it is important that there is an understanding that judges as 'guardians' of the justice system may play a significant role in relation to the judicial supervision, monitoring, engagement and development of online courts, and in creating clear boundaries in relation to Judge AI with an understanding that Judge AI will have differing evolutionary stages.

REGULATING AND HARMONIZING BIOMETRIC ECOSYSTEMS: ADDRESSING THE FULL SPECTRUM OF RISKS USING GLOBAL SAFETY MODELS AND CONTROLS

The spectrum of biometric technologies today is expansive, encompassing many types, or modalities, of biometrics, from face recognition to iris to fingerprint to DNA, used singly or in combination. Biometrics are a significant category of technology; at least 104 countries use fingerprint and/or iris as part of national ID systems,[**] and uses for biometrics extend well beyond the national ID system context in both private and public sector use.[***] Increasingly understood as a technology of concern based on scientific data, there is a growing body of law and regulations seeking to mitigate the risks associated with biometric technologies. Current approaches to biometric regulation thus far, however, have been insufficiently constructed, and do not address the full ecosystem of biometric modalities[****] and the risks that can arise from them.

Pam Dixon*

TPQ
Winter 2021/22

*Pam Dixon is an author, researcher, and the founder and Executive Director of the World Privacy Forum, a non-profit public interest group. She writes extensively about face recognition and biometrics, including in peer-reviewed studies.

**ID4D 2018 Global Dataset, 2020 updates. World Bank, (2020). https://datacatalog.worldbank.org/search/dataset/0040787

*** Uses range from authentication uses, such as face or fingerprint authentication for mobile phones, to uses in the criminal investigation context. Diverse commercial applications exist, for example, photo management applications that use face recognition, or biometric-based employee clock-in devices designed for workplaces.

**** Biometric modalities include: DNA, face, fingerprint, speech, voice, iris, retina, periocular, ear, gait, tattoo, heartbeat, hand geometry, odor, behavioral, typing recognition, vein recognition, for example. See Biometrics Institute: https://www.biometricsinstitute.org/what-is-biometrics/types-of-biometrics/

In recent years, biometrics have been advanced by deep learning architectures and techniques, and the advances far outpace policy adaptations.[1] The full biometric ecosystem – from data collection to algorithmic quality, from the hardware of biometric systems, to the implementation of the full systems – has components that create, or can create, meaningful risks. These risks have been rigorously documented and are no longer theoretical. The infrastructure components of face recognition systems, such as the cameras used, for example, can have a significant impact on the accuracy and functioning of the underlying algorithms.[2] In contactless fingerprint systems, which are still rapidly evolving, simple things such as how far away the fingers are to be scanned are making a difference in accuracy.[3] The primary controversies around face recognition systems[4] arise largely from the well-documented potential for racial, gender, and age[5] bias, as well as politically-driven utilization of such systems.[6] Additionally, some face recognition systems rely on unconsented data collections, which is also controversial.[7]

As compelling as it is to discuss biometric systems and risks in isolation, for example, by focusing on a single modality such as face recognition or DNA or even on a particular use case, it is essential to contextualize biometric systems in the broader context of all biometric modalities, and view biometrics as a complete ecosystem of multiple biometrics. Additionally, biometric modalities are often layered in what is called a multi-modal or multi-biometric approach, combining face

[1] Certain modalities of biometrics, particularly face recognition systems, have been profoundly changed by advances in certain AI architectures, specifically, Convolutional Neural Networks (CNNs). See, Mayank Vatsa, Richa Singh, Angshul Majumdar, Ed. *Deep Learning for Biometrics* (CRC Press, 2018).

[2] David Leslie, "Understanding bias in face recognition technologies, an explainer," *The Alan Turing Institute, October 2020.* https://www.turing.ac.uk/sites/default/files/2020-10/understanding_bias_in_facial_recognition_technology.pdf

[3] J. Libert, J. Grantham, B. Bandini, K. Ko, S. Orandi and C. Watson, *NIST Report (NISTIR) 8307: Interoperability Assessment 2019: Contactless-to-Contact Fingerprint Capture, NIST, (2019).* https://nvlpubs.nist.gov/nistpubs/ir/2020/NIST.IR.8307.pdf

[4] European Parliament Resolution A9-0232/2021: *Artificial Intelligence in Criminal Law and its Use by Police and Judicial Authorities in Criminal Matters, (2020/2016/INI), adopted 6 October 2021, Strasbourg, France.* https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_EN.html, sections 25-31 in particular.

[5] Age bias in face recognition occurs in both younger and older individuals. See Anil Jain, *Biometric Recognition of Children, Challenges and Opportunities* (Michigan: Michigan State University, 7 June 2016). http://biometrics.cse.msu.edu/Presentations/AnilJain_UIDAI_June7_2016.pdf. See also Patrick Grother, Mei Ngan, Kayee Hanaoka. *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects in Facial Systems*, NIST (December 2019), https://nvlpubs.nist.gov/nistpubs/ir/2019/ NIST.IR.8280.pdf.

[6] Patrick Grother, Mei Ngan, Kayee Hanaoka. *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects in Facial Systems*, NIST (December 2019), https://nvlpubs.nist.gov/nistpubs/ir/2019/ NIST.IR.8280.pdf. Regarding political risks, see: Teki Falconer and S. Okedara, "National Security Exemptions, The Dark Side of Identity Systems," Episode 24, Segment 3. https://id4africa.com/livecast-ep24-the-dark-side-of-identity-part-2/

[7] *Facial Recognition: the CNIL orders Clearview AI to stop reusing photographs available on the Internet,* CNIL France, 16 December 2021. https://www.cnil.fr/en/facial-recognition-cnil-orders-clearview-ai-stop-reusing-photographs-available-internet

and iris, or iris and fingerprint, and so on.[8] As discussed, current legal frameworks do not yet effectively address the full range of biometric modalities and the risks associated with them, resulting in incomplete policy protections and fragmentation. Fragmented approaches toward the regulation of biometrics is a category risk in and of itself, and ultimately, fragmented approaches are not a sustainable solution for the types of meaningful risks biometrics can pose within jurisdictions, and across jurisdictional boundaries. Even the most robust policy solutions implemented thus far to mitigate the risks of face recognition systems are extremely limited when compared to the full continuum of biometric modalities that require regulatory attention.

"Further, the policy tools that have been the primary focus thus far for biometric controls are important, but nonetheless incomplete for the job; these include principles for responsible use, the utilization of consent mechanisms as a control, and bans or moratoriums."

Currently, more than 145 jurisdictions have passed comprehensive national-level data governance and protection frameworks.[9] Because these frameworks often use generalized language regarding protections for biometric data, the existing rules cover biometrics, but often lack specifics tied to any particular modality of biometric. In the European Union (EU), the EU General Data Protection Regulation (GDPR),[10] covers biometrics as a sensitive data category. The GDPR does not address specific face recognition, iris, or multi-modal concerns, and does not address face recognition uses for some important use cases, for example, national security uses of biometrics are not covered under the GDPR. Beyond the GDPR, in multiple jurisdictions, regulatory solutions for biometric risks often have a strong emphasis on the modality of face recognition, iris, fingerprint, or DNA, and do not reflect a more mature legislative model.[11] The current approaches to biometric regulation are well-meaning but incomplete, and have created meaningful gaps in protections. On one hand, comprehensive legislation is too broad to be specific enough. On the other, a focus on single-modality biometric legislation is too narrow, and leaves gaps

[8] Multimodal biometrics and biometric fusion are instances when one or more biometric attributes are used together. The term *multibiometric may also be used. See:* Maneet Singh, Richa Singh, Arun Ross. "A comprehensive overview of biometric fusion," *Information Fusion,* Vol. 52 (2019): p. 187-205. https://www.cse.msu.edu/~rossarun/pubs/SinghRossBiometricFusion_INFFUS2019.pdf

[9] Graham Greenleaf, "Global Data Privacy Laws 2021: Uncertain Paths for International Standards," (11 February 2021), p. 169; *Privacy Laws & Business International Report*, p. 23-27, Available at SSRN: https://ssrn.com/abstract=3836408

[10] *EU General Data Protection Regulation*, http://www.privacy-regulation.eu/en/index.htm. The GDPR went into effect 25 May 2018. See in particular Article 9.4.

[11] *Biometric Information Protection Act (760 ILCS 14) (Illinois)* https://ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004

regarding the other modalities, and gaps regarding multi-modal systems.

Further, the policy tools that have been the primary focus thus far for biometric controls are important, but nonetheless incomplete for the job; these include principles for responsible use, the utilization of consent mechanisms as a control, and bans or moratoriums. For example, subnational legislation in the United States tends to focus on single biometric modalities, such as face recognition, and consent is often the primary tool utilized for protection.[12] Consent, when utilized by itself without any other administrative or procedural controls or underlying protections, provides quite poor protections in the biometric context.[13] The failure to conceptualize biometric guardrails in a more sophisticated ecosystem approach will lock in fragmented approaches, which over time, promises to create regulatory havoc and gaps in protections. It is an unsustainable policy strategy for national or sub-national legislatures to craft multiple, stand-alone, and possibly rivalrous biometric policies for separate biometric modalities and selected biometric use cases.

The chemical safety model, because it is both broad and granular, reduces policy fragmentation with meaningful, measurable, and well-understood and documented risk evaluations and mitigations. The global body of chemical safety regulations exists in most countries to monitor chemicals that pose dangers to people and to protect people from a complex range of chemicals.[14] Chemical safety policies that are crafted at the national and subnational levels are built according to a common framework, use the same definitions, are fit for each jurisdiction, and are also harmonized globally while respecting jurisdictional contexts. Biometrics, as an equally complex data ecosystem with overarching risks and particular risks attached to specific modalities, could be similarly regulated under an umbrella of controls derived from safety and risk-oriented regulatory models, with each biometric modality receiving appropriate and granular regulatory attention befitting the modality being considered. Additional risks that arise from mandatory biometric enrolment in some systems is an additional type of risk to consider and mitigate, among other types of risks.

---

[12] *Biometric Information Protection Act (760 ILCS 14) (Illinois)* https://ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004, Capture or use of Biometric Identifiers, Sec. 503.001. Title 11. (Texas) https://statutes.capitol.texas.gov/Docs/BC/htm/BC.503.htm

[13] Consent has important uses, particularly in human subject research, where consent must be meaningful. International norms have developed around this context for consent. See: 21 CFR 50.20 *General Requirements for Informed Consent. See also: 45 CFR part 46 and HHS, Federal Policy for the Protections of Human Subjects ('Common Rule')* https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html. Consent as a primary tool for all data protection and privacy contexts, however, has been replaced by policies around legitimate basis for processing, among other protections. See "Click to Consent? Not good enough anymore," *Privacy Commissioner of New Zealand, 2019.* https://privacy.org.nz/blog/click-to-consent-not-good-enough-anymore/

[14] United Nations, *GHS*, https://www.unece.org/trans/danger/publi/ghs/ghs_welcome_e.html; The World Health Organization operates the International Programme on Chemical Safety (IPCS), https://www.who.int/health-topics/chemical-safety#tab=tab_1

The most salient hallmark of chemical safety regulation models is the robust procedural and administrative controls they use, joined with highly granular applicability. For example, the regulations present an umbrella under which many types of chemicals can be individually regulated. Lead and arsenic, for example, are regulated under the same overarching framework, but lead and arsenic have differing protections and cutoff points, as is appropriate given the toxicological differences. In chemical safety regulations, the country-level legal frameworks are then harmonized by two multilateral institutions, the World Health Organization and the United Nations. The UN has a program called the *Globally Harmonized System of Classification and Labelling of Chemicals* (GHS), which is regularly updated.[15] The idea of the UN GHS is to bring a global, standardized approach to chemical safety across all jurisdictions.[16] Labeling is to be the same, level or grade of the risk is the same, and risk mitigation strategies would be similarly harmonized internationally. The UN GHS plan is part of the implementation of the Sustainable Development Goals (SDGs).

"These administrative and procedural controls provide a toolbox of options that go beyond best practices, simple consent structures, and narrow bans, and could readily be utilized in biometrics. Under this type of framework, all biometric modalities would be regulated under one umbrella."

The key controls in chemical safety models include:

- Pre-market safety, quality, and other risk assessments and requirements
- Registration of products
- Ongoing product documentation
- Audits
- Post-implementation surveillance (observation) and documentation
- Compliance labeling

[15] *GHS, Rev. 8, (2019). United Nations.* https://unece.org/ghs-rev8-2019

[16] United Nations, *GHS*, https://www.unece.org/trans/danger/publi/ghs/ghs_welcome_e.html

- Safety certifications
- Technological proof of compliance and risk mitigation
- Ongoing review, oversight, and multistakeholder feedback (and complaint mechanisms).

These administrative and procedural controls provide a toolbox of options that go beyond best practices, simple consent structures, and narrow bans, and could readily be utilized in biometrics. Under this type of framework, all biometric modalities would be regulated under one umbrella. Each modality would be subject to meaningful administrative and procedural controls. All biometrics -- DNA and face recognition, along with the other biometric modalities -- would be evaluated under the auspices of the regulation, with their own measures for accuracy and pre-market fitness.

In practice, this would mean that before a biometric product could be put out in the market, it would have to be assessed for pre-market safety, quality, and other risks. For face recognition systems, this would mean that the product could not be discriminatory in its operations, and the risk points would have rigorous testing. For example, age, gender, and race biases would be tested. Each biometric product, after passing the assessment, would be registered, labelled, and would be required to submit documentation to regulators. Regulators would be able to conduct audits, and there would be a post-implementation market surveillance and documentation program. Safety certifications would need to be met, and biometric products would need to proactively provide proof of compliance and mitigate known risks. And finally, there would be an ongoing review of the biometric products, consumer and end-user feedback, as well as formal complaint mechanisms.

In considering how this system could be worked out in practice, it is useful to review some exemplar implementations regarding chemical safety regulatory models. The regulations are notable, in that they are harmonized across borders, yet the regulations are also fit for each country-level context of economic and technological development. This is important and would need to be present for any harmonized biometric approaches.

The EU has two significant member state-wide regulatory models in chemical safety. Both regulations offer excellent tools for mitigating harms. REACH[17] is the European Regulation on Registration, Evaluation, Authorization and Restriction of Chemicals. It entered into force in 2007, replacing the former legislative framework for chemicals in the EU. This important and precedent-setting regulation applies

[17] European Commission, *REACH*, https://ec.europa.eu/growth/sectors/chemicals/reach_en

to essentially every chemical product manufactured, imported, or sold within the EU. Manufacturers and importers must register all substances produced above a set yearly volume, and must identify risks associated with the substances they produce, demonstrate compliance in mitigating the risks, and establish safe use guidelines for the product so that the use of the substance does not pose a health threat.

Another precedent-setting regulation, RoHS,[18] applies to any business that sells electrical or electronic products, equipment, sub-assemblies, cables, components, or spare parts directly to RoHS-directed countries. Products must be cleared for market prior to launch. All parties in the supply chain must provide documentation/ recordkeeping, regularly update information, mandatory compliance labeling. All of these features could be helpful in regulating biometric products. Other countries that have enacted RoHS include Japan, Korea, and China. In the U.S., the states of California, Colorado, Illinois, Indiana, Minnesota, New Mexico, New York, Rhode Island, and Wisconsin, among others, have enacted RoHS-like and e-waste regulations.

In the U.S., a federal statute, the Chemical Safety for the 21st Century Act,[19] regulates chemical substances of concern. The statute has these compliance requirements: pre-manufacture notification for new chemical substances prior to manufacture, where risks are found (after risk assessment), testing by manufacturers, importers, and processors, certification compliance, reporting and record keeping. If a substance presents a substantial risk of injury to health or the environment, the party must immediately inform the EPA. As mentioned earlier, in addition to the Chemical Safety for the 21st Century Act, some states adopted additional EU-style regulations after the European RoHS model.

Most countries in Africa already have regulations in place that assert legally binding controls on toxic substances. Lead is one example of a regulated toxic substance under a variety of such laws in African countries. In Algeria, for example, Arrêté No. 004/MINEPDED/CAB of 21 September 2017, modifies and completes the list of chemicals in Décret No. 2011/2581/PM of 23 August 2011, which regulates dangerous chemicals. Among other controls, the regulations prohibit the manufacture, sale and import of paints containing more than 90 ppm of lead (10/8/17). Algeria, Cameroon, Ethiopia, Kenya, South Africa, and Tanzania are among the African countries that have similar regulations.

India has adopted the National Action Plan for Chemicals; India's version of

[18] European Commission*, RoHs Directive*, Current: (2011/ 65/ EU). First RoHS Directive: (2002/95/EC), https://eur-lex.europa.eu/legal-content/en/TXT/?uri=celex%3A32011L0065

[19] U.S. Environmental Protection Agency, *Chemical Safety for the 21st Century Act*, https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/frank-r-lautenberg-chemical-safety-21st-century-act

safety regulations for hazardous chemicals.[20] In the past, India's regulations were not modeled after "REACH," the EU regulation. In late 2019 and continuing into 2020, however, India embarked on the creation of a National Action Plan for Chemicals (NAPC) to move into a more REACH-like system.[21] The idea is to create a harmonized system of classification of toxic chemicals that complies with the UN's Global Harmonization Strategy for chemical safety. Helping this effort is India's standing committee for chemical safety legislation, the National Coordination Committee (NCC) under the Ministry of Environment, Forest and Climate Change (MoEF&CC).[22] The late 2019 draft National Action Plan for chemical safety for India recommendations to compile a national chemicals inventory; analyze and assess the risks of those chemicals; implement the UN Global Harmonization Strategy (GHS); and develop risk mitigation strategies, policies and regulations.

The rich, adaptable, comprehensive, and yet granular chemical safety models in use today provide a pathway to move biometric regulation away from the fragmented and ineffective current practices that rely on consent, bans, and single-modality regulations. It is essential that complex biometric ecosystems are regulated consistently, comprehensively, and in a manner that is harmonized across jurisdictions so that biometrics are not subject to fragmentation that results in inconsistent or weak data governance, security, and privacy protections. Biometrics, as a technology of concern, merits high levels of attention to administrative and procedural controls, as well as a focus on harmonization on key aspects of regulation, such as agreement on definitions.

[20] National Disaster Management Authority of India, *Chemical Disaster Page*, https://ndma.gov.in/en/2013-05-03-08-06-02/disaster/man-made-disaster/chemical.html

[21] *Chemical Watch*, "India's draft national plan includes inventory and registration," 6 January 2020, https://chemicalwatch.com/86343/indias-draft-national-chemical-plan-includes-inventory-and-registration

[22] Government of India, Ministry of Environment, Forest, and Climate Change, http://moef.gov.in

GOVERNMENTS AS REGULATORS AND CONSUMERS OF ETHICAL AI

Algorithmic Decision-Making (ADM) and Digital Technologies have created complex and multifaceted challenges for governments and the delivery of public services. Governments must operate simultaneously as the regulator and as a consumer, while maintaining public trust. The roll out of ADM systems has provided significant failures in public policy. For example, the Federal Government of Australia and the New South Wales State Government both deploying systems with significant flaws which led to serious outcomes for both welfare recipients and governments just in the last 10 years. Solutions to confront the failure of ADM systems can be addressed more effectively at a procurement and contract level as well as through the accountability and understanding within the public service. In order to ensure the fair use of digital technologies, governments must be willing to scrutinize their own purchase of technology, ideally through the lens of AI ethical frameworks, and address their own governance and procurement policies, thereby fulfilling their responsibility as service providers to citizens.

Gabby Bush* & Jeannie Paterson**

TPQ
Winter 2021/22

* Gabby Bush is the Program Manager at the Centre for AI and Digital Ethics at the University of Melbourne.
** Jeannie Paterson is a Professor of Law and the Co-Director of the Centre for AI and Digital Ethics at the University of Melbourne.

Increasingly sophisticated digital technologies provide significant opportunities for governments in making consistent, informed and efficient decisions. Thus, big data, algorithms, expert systems and machine learning are increasingly used to make predictions, recommendations and even decisions about access to government services, a process sometimes known as Algorithmic Decision-Making (ADM). While the safety, equity and fairness of these systems and tools are often criticized, governments continue to pursue the use of ADM in public service delivery. In this process, governments have consistently contracted out the development of ADM systems to private companies, while continuing to face significant backlash when these systems fail.[1] Simultaneously, as digital technologies increasingly present the risk of harm to many parts of society, governments are endeavoring to find solutions for protecting citizens when using technology, including through regulatory and legal responses.[2] Existing in this new technological landscape as regulator and legislator, consumer and contractor, is a unique challenge for governments.

Our contention is that to pursue safer and more ethical uses of digital technologies and AI across society, governments must look inwards and scrutinize their own processes in both contracting for and relying on ADM. ADM as an alternative to human processing and in augmenting human decision-making is not neutral: it has the potential to replicate and amplify existing biases, while also making the basis for decisions harder to interrogate. Significantly, ADM algorithms relied on by governments are built by private sector entities. In using outsourced ADM systems, governments risk letting decisions of the democratic state becoming decisions of the private sector.

In the last five years, Australia has utilized ADM algorithms to recover debt from welfare recipients that have perpetuated errors, including removing large amounts of money from the bank accounts of vulnerable people, and ultimately have been found to be outside the scope of the discretionary decision making power available to the cognate agencies. Governments in the U.S. and in the EU have used ADM systems for other purposes, with similar outcomes.[3] There are, however, better options for procurement and use of ethical AI in governments. Tighter

[1] Jon Porter, "UK ditches exam results generated by biased algorithm after student protests," *The Verge*, 17 August 2020. https://www.theverge.com/2020/8/17/21372045/uk-a-level-results-algorithm-biased-coronavirus-covid-19-pandemic-university-applications; Rebecca Turner, "Robo-Debt condemned as a 'shameful chapter' in withering assessment by federal court judge," *ABC,* 11 June 2021. https://www.abc.net.au/news/2021-06-11/robodebt-condemned-by-federal-court-judge-as-shameful-chapter/100207674

[2] See eg JM Paterson, S. Chang, M. Cheong, C. Culnane, S. Dreyfus and D. McKay, "The Hidden Harms of Targeted Advertising by Algorithm and Interventions from the Consumer Protection Toolkit." *International Journal on Consumer Law and Practice*, Vol. 9 (2021), p. 1-24.

[3] "Proposal for a Regulation of the European Parliament and of the Council (EU) COM (2021) 206 final laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts [2021]," (EU Draft AI Act.) Australian Human Rights Commission, *Human Rights and Technology Final Report* (2021).

regulation and more robust governance systems, teamed with ethical AI standards and scrutiny in AI procurement contracts, offer practical ways for governments to use new digital technologies to support rather than erode democratic values. Once governments have enacted fairer, more effective processes for the contracting and deployment of technology within their own services, these insights should flow through to governments' regulatory strategies for technology in society. Governments gain regulatory crediablity for having their own house in order.

"While there are potential benefits in using new technologies, and particularly the insights from data, to government functions, those benefits lie in specific discrete field, perhaps autonating clerical task or providing empirical evidence to guide policy."

Algorithmic Decision-Making and Streamlined Services

Governments are often interested in streamlining processes and optimizing public service delivery. Ongoing cost cutting is a common and necessary process as funds must be divided amongst many portfolios. A body of literature points to the opportunities that AI technologies provide governments for more efficient processes.[4] There are significant benefits and incentives for using automation – the utilization of available data allows for evidence based decision-making. Automation allows governments to fast track routine tasks that do not require discretion or judgement. Moreover, newer digital or algorithmic technologies, such as those premised on machine learning, computer vision, natural language processing or 'artifical intelligence', not only allow processing large amounts of data but allow opportunities to use that data to inform decision making in an adaptive and dynamic way.[5] In the Australian Government, the Digital Transformation Agency and several statutes encourage the use of technology throughout government including for the automation of decisions and processes. [6] Digital technologies can be innovative and novel in the functioning of the state, causing net benefit. For example, computer vision has been used to pick up illegal mobile phone use by drivers in New South Wales with the aim of reducing road accidents.[7]

[4] Rajan Gupta and Saibal Kumar, *Introduction to Algorithmic Government* (Palgrave Macmillan, 2021), p. 5; Michael Veale and Irina Brass, "Administration by Algorithm?," in *Algorithmic Regulation* (Oxford University Press, 2019), edited by Karen Yeung and Martin Lodge, p. 121 – 142.

[5] Ben Dor, Lavi M and Cary Coglianese, "Procurement as AI Governance," *IEEE Transactions on Technology and Society, Technology and Society*, Vol. 2, No. 4 (2021): p. 192.

[6] Anna Huggins, "Addressing Disconnection: Automated Decision-Making, Administrative Law and Regulatory Reform," *University of New South Wales Law Journal,* Vol. 44, No. 3 (2021): p. 1059.

[7] New South Wales Ombudsman, *The New Machinery of Government: Using Maching Technology in Administrative Decision-Making*, Special report under section 31 of the *Ombudsman Act 1974,* 29 November 2021.

While there are potential benefits in using new technologies, and particarly the insights from data, to government functions, those benefits lie in a specific discrete field, perhaps autonating clerical task or providing empirical evidence to guide policy. Once the ADM systems are used to make *discretionary* decisions, such as the allocation of welfare, the focus of policing, whether children are at risk or the likelihood of adolescent recidivism, we need to ask whether this kind of use of technology is consistent with the aspirations of good governance.

Flaws and irregularities in the performance of government functions may be amplified by ADM.[8] Algorithmic decision-making systems can perpetuate a toxic status quo, entrench bias, discriminate against the marginalized, and lack human-based judgment and reason which cannot be fully quantified or written into code.[9] Rubenstein wrote on this that "AI systems are social artifacts that embed and project human choices, biases and values."[10] Typically the justification for using these technologies is that using algorithmic processes removes the human bias, creating more efficient, neutral and replaceable decisions. Yet, critics argue that "Individual administrators can, and do, make errors but not at the scale and speed of ADM."[11]

The Risks of ADM in Government

Using ADM systems in government services carries considerable risks of harm for citizens and also the government's social license. The primary risk is to citizens who may be the recipient of poor or biased decision-making at the hands of an algorithm – they may be the recipient of an incorrect debt notice, considered by an algorithm to be at risk of recidivism due to their race or denied a loan due to their previous postcode.[12] The deployment of flawed and biased systems undermines a core function of democratic governments, which is to treat its citizens fairly and respectfully, as well as investing public funds prudently. As seen in the Australian experience, poorly designed and deployed algorithmic systems are ultimately vulnerable to challenge by those institutions that oversee government exercises of power, such as tribunals, courts and ombudsmen. This may mean that government investment in algorithmic processes is ultimately rendered futile, as well as causing considerable harm to the typically already vulnerable individuals affected by the process.

The reasoning for increasing government use of ADM is complex – at best, the algorithms are used in good faith, to conserve public funds and streamline services;

[8] New South Wales Ombudsman, 29 November 2021.

[9] Veale and Brass, (2019), p. 122; Virginia Eubanks, *Automating Inequality : How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's Press, 2018); NSW Ombudsman, "The new machinery of Government,".

[10] David S. Rubenstein, "Acquiring Ethical AI," *Florida Law Review,* Vol. 73, No. 4 (2021): p. 761.

[11] Huggins, (2021), p. 1058.

[12] Daniel Ziffer, "Threat of 'Postcode Discrimination' as credit scores skewed by where you live," *ABC,* 7 February 2022. https://www.abc.net.au/news/2022-02-07/threat-of-postcode-discrimination-in-credit-scores/100723574

at worst, the algorithms are used by governments looking to restrict access to social services in a manner which removes direct accountability from the public service in their deployment. Virginia Eubanks most famously referred to ADM in her book *Automating Inequality* as 'systems which punish the poor.'[13]

> *"Ethical AI for procurement contracts would indeed address the issues of AI for government systems. However, this approach needs to be multifaceted."*

Regulating Digital Technologies

Governments are increasingly under pressure to increase regulation surrounding the use, consumption and creation of emerging digital technologies and to address ongoing issues of misinformation, deep fakes and the erosion of privacy online. The regulatory landscape is complex, with the digital platforms responsible for developing many of the new uses of AI transcending traditional nation-state boundaries. However, significant concerns highlight the need for this regulation, including a focus on uses of AI in ADM. Currently, many frameworks for the ethical use of AI have been written at a national and multinational level. These frameworks are typically not legally binding: in most cases they are at most a form of 'soft law' influencing behaviour without mandatory force. Increasingly, however, efforts are being made to formalize laws about ADM and AI.[14]

Governments are responsible for these regulatory systems. They are also effectively consumers in their own right. Governments have always held large amounts of data on their citizens and are now using ADM to process that data and deliver public services. In doing so, governments are faced with the challenge of not only trying to understand how to regulate the industry effectively for the safety of its citizens, but also how to operate and procure these systems as a purchaser. In this capacity, governments in some broad ways face challenges usually associated with traditional consumers, namely of considerable information assymetries vis-à-vis the companies offering the technology. These asymmetries impinge on governments' ability to contract for optional outcomes in technological tools. The added factor is that governments' decisions about their consumption of technology impact on the individuals or citizens they are supposed to represent.

[13] Eubanks, (2018).

[14] Australian Human Rights Commission, (2021).

Private Companies and the State

Governments are usually not equipped to develop their own software, so the development of this software is outsourced to private companies. These companies are driven by commercial objectives, including, significantly, safeguarding their intellectual property and trade secrets.[15]

With ADM, programmers need to decide the outcomes, weight decisions and prioritize information. The concern for automating government services is that the decisions once made by politicians or bureaucrats become decisions made by private companies. As Rubenstein writes "the government's decision to *outsource* AI development or deployment can be highly consequential. Especially if private vendors become de facto deciders."[16] Where governments have systems in place for transparency and accountability, private companies are required to do no such thing. Companies moreover tend to try to protect their intellectual property and operate with less than full transparency. This is cause for serious concern as private companies have increasing input into the decisions made and enacted by the state. The further concern is that the private companies are operating on a for-profit model.[17] Are decisions made by a private, for-profit entity the best decisions for citizens of a democratic state?

Failures of ADM Systems

There have been well-documented examples of ADM in government services which have resulted in problematic outcomes, and ultimately the withdrawal of the system, including in Australia, England, the EU[18] and the U.S.[19] In 2020 the UK Government, from the Office of Qualifications and Examinations Regulation used a grading algorithm to decide the outcomes of students' A-Levels which had been disrupted by the COVID-19 Pandemic.[20] This algorithm used factors to weight the score, including the historical scores of the school each student attended. This meant that high-performing students from low-performing schools were given lower grades, while those who attended private schools were graded on average at a much higher level by the algorithm. This ADM system received significant backlash and with no processes for contesting the outcome the algorithm was pulled and the withdrawal

[15] Dor and Coglianese, (2021), p. 193.

[16] Rubenstein, (2021), p. 768.

[17] Sarah Valentine, "Impoverished Algorithms: Misguided Governments, Flawed Technologies, and Social Control," *Fordham Urban Law Journal*, Vol. 46, No. 2 (2019): p. 364–427.

[18] Australian Human Rights Commission, (2021).

[19] See generally Eubanks, (2018).

[20] Gabby Bush, Henrietta Lyons and Tim Miller, "Data isn't neutral and neither are decision algorithms," *Pursuit,* 15 September 2020. https://pursuit.unimelb.edu.au/articles/data-isn-t-neutral-and-neither-are-decision-algorithms

of the algorithmic grades completely.[21]

The Federal Government of Australia deployed an algorithm now known as Robo-Debt, to issue debt notices to welfare recipients, from 2016.[22] The algorithm calculated the amount of welfare received and averaged this sum against the reported income of that year for each recipient. This algorithm found that thousands of welfare recipients had been overpaid and issued debt notices to recover the overpayments. However, the calculations made by the algorithm were in many cases, incorrect. For example, the algorithm calculated the average of a yearly income – not taking into account fluxuations of income and time spent on Centrelink. Nor did the algorithm account for "working credits" a scheme which allows welfare recipients to work for an allotted amount of hours without losing their welfare.[23] The result was that some 400,000 Centrelink recipients received false debt notices, totaling $720 million and $400 million in unlawful demands.[24] There was then no avenue for contesting this decision and in some cases, these debt notices were attributed as the cause of suicide in recipients.

This was not the last time that a government in Australia has used an algorithm to recoup debt from welfare recipients. Between 2016 and 2019 the New South Wales Government used a similar algorithm to garnish overpayments from welfare recipients.[25] While the algorithm was similar to the robot-debt algorithm, this one went one step further – instead of issuing a debt notice to the recipient in question, the algorithm gained access to the bank account of the recipient and removed the supposed debt directly from their account.

In 2021 the NSW Ombudsman released a report detailing the failure of this policy and the steps that should be taken to avoid a repeat of this issue. Pertinently, the report noted that in Australia there is no clear and available information on the extent of use of ADM in government decision-making. The report detailed a set of standards for the authorization of machine technology at a parliamentary level. These standards included a series of properties:

[21] Catherine Caroll-Meehan, "A-levels governments u-turn has left universities in the lurch," *The Conversation,* 20 August 2020. https://theconversation.com/a-levels-governments-u-turn-has-left-universities-in-the-lurch-144763

[22] Luke Henriques-Gomes, "Robodebt: government to refund 470,000 unlawful Centrelink debts worth $721m." *The Guardian*, 29 May 2020. https://www.theguardian.com/australia-news/2020/may/29/robodebt-government-to-repay-470000-unlawful-centrelink-debts-worth-721m

[23] NSW Ombudsman, (2021), p. 27.

[24] Luke Henriques-Gomes, "Robodebt class action coalition agrees to pay 12bn to settle law suit," *The Guardian*, 16 November 2020. https://www.theguardian.com/australia-news/2020/nov/16/robodebt-class-action-coalition-agrees-to-pay-12bn-to-settle-lawsuit

[25] NSW Ombudsman, (2021), p. 11.

- Is it visible?
- Is it avoidable?
- Is it subject to testing?
- Is it explainable?
- Is it accurate?
- Is it subject to audit?
- Is it replicable?
- Is it internally reviewable?
- Is it externally reviewable?
- Is it compensable?
- Is it privacy protected and data secure?

The NSW Ombudsman found that while it was permitted by legislation to issue garnishee orders, Revenue NSW had adopted new standards for garnishing debt, including a 'minimum protected balance,' meaning that the garnisheeing of debt was not allowed when the debtor had under a certain amount of funds.[26]

These failures were ultimately addressed through litigation in court and review by the Ombudsman. Systems that were adopted to streamline processes had to be remedied through the legal system and independent bodies. The result was not more efficient systems for government. In other words, without proper procurement and governance processes, there is a false efficiency. ADM systems do not result in better outcomes for both citizens and government.

Building Accountability

The Australian cases raise a multitude of questions about government use of ADM, and also about the outsourcing the build of technology that impacts directly on core government functions, and, inevitably, vulnerable populations.[27] They signal the need for serious consideration of the governance of ADM systems in government, and also the processes by which they are contracted, deployed, scrutinized and monitored by governments.

Procurement Contracts for ADM in Government

As already noted, ADM systems are commonly not developed by governments themselves. Instead they are often, although not always, procured by government contracts to be delivered by private sector entities. Ben Dor and Coglianese in 2021, along with David S. Rubenstein have advocated for an approach that would see

[26] NSW Ombudsman, (2021).

[27] Valentine, (2019), p. 368.

ethical AI standards for accountability and transparency built into procurement contracts for government AI systems.[28] In doing so, the contracts would be reviewed and re-written for each time new technology is procured. The research claims that through this procurement process, ethical standards for ADM systems would be constantly updated through the process of contracting and procurement as new technology and AI systems are developed.[29] In principle, this approach would provide immediate solutions for ethical AI in government, however it falls heavily on those in the public service procurement departments to ensure all procurement contracts contain up to date ethical frameworks to ensure the ethical AI requirements are standardized and contracts are delivered adequately.

This approach has merits, as most government services are not developing their own AI systems it makes logical sense that the technology provided through procurement, be subject to ethical requirements in the contract.

Dickenson-Delaporte et al, in their work on self-regulation for the advertising sector noted that self-regulation would require action on behalf of the firm in a global market where the government will not step in.[30] The accountability placed on the private sector through procurement contracts, again removes the responsibility of the government and instead requires the private sector to prove their ethical practices but not vice versa. Governments need contracts to ensure they can scrutinize the ethical practices. As Rubenstein aptly wrote "some AI models *cannot* be explained because of their complexity, other AI models *will not* be explained because they don't have to be."[31] The role of the contract is to ensure that the ADM System has to be explained both to and by the government.

Towards Ethical AI Procurement

Ethical AI for procurement contracts would indeed address the issues of AI for government systems. However, this approach needs to be multifaceted. First there must be regulation *requiring* this practice and oversight into these systems. Including ethical clauses in procurement contracts should not be undertaken at the discretion for the contracting agency or department.

Secondly, the procurement contracts should provide the framework for government scrutiny over ADM systems. Rather than the onus being placed on the contractor,

[28] Ben Dor et al, (2021), p. 192.

[29] Rubenstein, (2021), p. 779; Ben Dor et al, (2021), p. 192.

[30] Sonia Dickinson-Delaporte, Kathleen Mortimer, Gayle Kerr, David S Waller and Alice Kendrick, "Power and responsibility: Advertising self-regulation and consumer protection in a digital world," *Journal of Consumer Affairs,* Vol. 54, No. 2, (2020): p. 675–700.

[31] Rubenstein, (2021), p. 779.

by the contract; it must instead be agreed that the ethical clauses give government oversight – they are able to scrutinize the design and development of the algorithm, they are able to revisit the system to understand the outcomes that are produced and the government has oversight on any changes made to the algorithm from inception to delivery and maintenance.

Thirdly, in order for this to generate successfully enabling environments, it must be created within the government service to ensure the ethical delivery of ADM. It is not enough to outsource the build of an ADM system to a private company – departments looking to do so must also have computer science and machine learning expertise within their department. In order to properly scrutinize the system, governments must have the personnel to understand that system in its entirety.

Governments must also ensure there is a team able to respond to complaints about the system. It is not enough to simply deploy an algorithm nor is it enough to have one technical expert responsible for all queries and concerns. Government agencies or departments relying on ADM systems should ensure there are systems and *processes* for citizens and affected individuals to contest the operation of those built into the very way in which they use ADM.[32] Humans must work in *collaboration* with the ADM system, acknowledging that the machine is not a flawless decision-maker.

Concluding Remarks

ADM systems in government provide a complex set of challenges that require detailed and multifaceted approaches. The notable failures we have highlighted primarily in Australia, require a significant overhaul and cultural shift in governance around ADM systems. In this piece we have focused on the procurement question. Where the decision is made to use an ADM system, any contracting out to a private company of responsibility for the design, construction and deployment of an ADM system must itself be subject to scrutiny and robust governance mechanisms by the government.

As both regulators and consumers, governments exist in a conflicting and unique position. It is not enough to explore regulation for the private sector, governments must also scrutinize their own processes. To fulfill their responsibility to the public, and their role as regulator, governments regulator and legislator, consumer and contractor, must consider the structural need for ethical ADM contracting and fairer ADM systems as a matter of urgency. The fair and robust function of public institutions is essential to a democratic state and one that must not be compromised by a superficial response to the perceived opportunities that ADM provides.

[32] See further, Lyons Henrietta, Eduardo Velloso and Tim Miller, "Conceptualising Contestability," *Proceedings of the ACM on Human-Computer Interaction*, Vol. 5, No. CSCW1 (2021): 1-25. doi:10.1145/3449180

DEMOCRATIZING AI

How can we locate the position of Artificial Intelligence when it comes to democracy? Rather than trying to put a framework built on an agent-based or globalization contexed approach, this paper aims to evaluate the Artificial Intelligence from a contextual perspective. While doing so an attempt on understanding how it can function within a community gains a considerable foothold. To achieve a success in such an attempt, two of the most important democratic discourses that resonate on this ground emerge critical: rule of law and the sustainable development goals. Especially, the recent Covid pandemic gave a considerable case study through which we can completely comprehend both of the matters of how AI can be used in surveillance and how can we be sure that it can be democratized.

Mark Findlay*

TPQ
Winter 2021/22

* Mark Findlay is the Director, Centre for AI and Data Governance, SMU. This paper was greatly assisted by the additions, insights and criticisms by Jane Loo, Ong Li Min and Zhang Wenxi.

Introduction

Analytical Assumptions – Communitarian AI governance

It is easy to be critical of unspecified, or empirically untested relationships between AI and community,[1] and that is why in the context of trusted robotics we argue for interrogating trust reciprocation from human recipients to robots providing identified and community-prioritized functions.[2] More interestingly for the purposes of this paper is the possibility of active social relationships that give substance to AI and community offering the possibility of 'democratizing' AI governance, which will be touched upon in the conclusion.

To date, the *AI in community* language concerned with how community-mindful AI deployment can initiate relationships of trust within recipient communities has not progressed to a sufficiently developed considerations of governance and communitarian democracy beyond concerns that trust has regulatory impacts in such directions.[3] At least at the level of suggestion, this paper seeks to extend that thinking in arguing that through more equitable and inclusive deployment of AI (which considers the benefits to social infrastructure of technology advancement) the legitimacy of AI and big data use is enhanced for social as well as market purposes. This suggestion also approaches our interpretation of the role of empowered data subjects in digital self-determination (DSD), which unfortunately there is no space to detail in what follows.[4]

In recalling a brief case-study on AI-assisted pandemic surveillance[5] and exposing it to an additional analysis in terms of equality and inclusion (particularly locating on discrimination and vulnerability), it is possible to shift the democratic focus to the powers behind deployment, rather than on the technology itself. It would show how the positive benefits of this surveillance was extensively denied in the South World (absent inclusion). In this example, it becomes apparent how the selective and myopic application of technologized surveillance regimes (or their absence) over already discriminated groups exacerbates structural inequalities.[6] More generally,

[1] Mark Findlay and Willow Wong, "Kampong Ethics," *Artificial Intelligence & Society: Legal, Social & Developmental Narratives from Asia* (forthcoming).

[2] Wong Zhang and Findlay, "Trust and Robotics: A Human-centric Pathway to Robots in Community," (forthcoming).

[3] Mark Findlay and Willow Wong, "Trust and Regulation: An Analysis of Emotion," *CAIDG Research Paper*, (10 June 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3857447

[4] "Digital Self Determination," Centre for AI and Data Governance, n.d., https://caidg.smu.edu.sg/digital-self-determination

[5] This case-study can only here receive the most outlined attention. For the full scope of our work see - https://caidg.smu.edu.sg/covid-resources

[6] On how vulnerable groups are discriminated against through technologized surveillance and control measures during COVID-19 see: Mark Findlay et al., "The Vulnerability Project: The Impact of COVID-19 on Vulnerable Groups," *SSRN Electronic Journal*, (2021), https://doi.org/10.2139/ssrn.3907183

it will put forward how the legitimacy of surveillance policy and control powers at large have been damaged by inequitable and non-inclusive imposition, and created strains on governance and drifts away from democratic values.

> *"This paradox is one reason why AI in community is advocated as the governance location for AI and big data. Communitarian incarnations of democratic values are likely to be more closely aligned with and reflective of purposes for AI that are socially sustainable."*

In short, the thesis of the paper is that if AI in community is respectful of the democratic values[7] of inclusion, and equality (leading on to governance legitimacy fleshed out in the final sections), then the potential for AI to be better governed at the community level, and hereafter to support communitarian governance in wider applications, can be realized. The structure of the argument is to employ two important democratic discourses that resonate with AI governance in the community:

- rule of law and
- the sustainable development goals,

in critically reviewing how COVID control powers in terms of AI-assisted surveillance are said to challenge liberal democracy in a broad individual rights understanding, while leaving behind more long-lasting negative impacts on vulnerable communities. If these communities had been included in a more equal exercise of control powers, then the legitimacy of AI technology and its potential to legitimize democratic powers would be endorsed.

SDGs, Equality and COVID Surveillance and Control Policy – North/South Divide

The United Nations Sustainable Development Goals (SDGs) are premised on two grand but simple assumptions – for its diverse populations, that the globe is not equitable, and it should not be so. Equality of opportunity, equal access to the fundamentals of human dignity and equal protection under universal human rights are the essence of the UN mission. Translated into sustainable development, and

[7] Such 'respect' in terms of Digital Self-determination is not confined to a rights discourse which I question in any case. See Mark Findlay, "Flipping the Other Way: Access, Not Protection, and the Role of Rights," (16 April 2019). Available at SSRN: https://ssrn.com/abstract=3372677 or http://dx.doi.org/10.2139/ssrn.3372677

along with Goals 5 and 17 and their focus on gender equity, Goal 10 aims at reducing inequality within and among countries.[8] While the measures of inequality/equality are couched in variables that influence economic development aspirations, human equality can be considered as a fundamental democratic value – equality before the law.

Recently, there has been a suggested alignment between AI technological advancement and the achievement of the SDG's.[9] However, as some have cautioned, the global deployment of AI can be a hegemonic project[10] if it is driven by the economic motivations of big tech, located in a world trade model that has always been pitted to the disadvantage of vulnerable economies in the Global South.[11] Indeed, the *AI for social good* mantra is regularly absent a recognition of the profound and debilitating dependencies inherent in techno-colonialism.[12] This avoidance of the power imbalance that tech transformation can visit on vulnerable economies, enables the hope that AI global proliferation will necessarily achieve the sectoral power displacements essential for democratic sustainable development. However, putting power imbalance in the picture and instead AI may be a retarding agent for democratic values, unless by making clear the power imbalances this provides the opportunity for AI to correct these through power dispersal and thereby promote / catalyse democratic values.

Underlying a healthy circumspection about whether AI deployment across the globe will democratize in its wake, is the understanding that AI promotion by many state/corporate collaborations is seen as a silver bullet for stimulating economic growth, or at least avoiding no-growth/de-growth.[13] In any prevailing neo-liberal world trade model where AI is dis-embedded from public good priorities and captured by states and multi-national corporations, it is possible for AI deployment to reflect

---

[8] This SDG calls for reducing inequalities in income as well as those based on age, sex, disability, race, ethnicity, origin, religion or economic or other status within a country. See: "Goal 10: Reduce Inequality within and among Countries," United Nations - Department of Economic and Social Affairs Sustainable Development, n.d., 10, https://sdgs.un.org/goals/goal10

[9] Barbara Rosen Jacobson, "Digital Technology for the Sustainable Development Goals," *Diplomacy.Edu* (blog), n.d., https://www.diplomacy.edu/blog/digital-technology-sdgs/

[10] Nick Couldry and Ulises Ali Mejias, "The decolonial turn in data and technology research: What is at stake and where is it heading?," *Information, Communication & Society* (November 2021): p. 1–17. https://doi.org/10.1080/1369118X.2021.1986102;
Meredith Whittaker, "The steep cost of capture," *Interactions*, Vol. 28, No.6 (2021): p. 50–55. https://doi.org/10.1145/3488666

[11] Mark Findlay, "Property Resisted," in *Law's Regulatory Relevance: Property, Power and Market Economies* (Cheltenham: Edward Elgar, 2017): Chapter 5.

[12] Christopher Chase-Dunn, "The Effect of International Economic Dependence on Development and Inequality: A cross national study," *American Sociological Review*, Vol. 40, No. 6 (December 1975): p. 720-738

[13] "AI is the future of growth," *Accenture*, (2017), https://www.accenture.com/_acnmedia/pdf-57/accenture-ai-economic-growth-infographic.pdf; Jacques Bughin et al., "Notes from the AI frontier: Modelling the Impact of AI on the World Economy," Discussion Paper, *McKinsey Global Institute*, (18 September 2018), https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy#part1

and comply with the processes and institutions of democratic governance but not promote equality as it is determined in the SDGs.[14]

This paradox is one reason why *AI in the community* is advocated as the governance location for AI and big data. Communitarian incarnations of democratic values are likely to be more closely aligned with and reflective of purposes for AI that are socially sustainable. Polanyi argues that in exchange market economies, the further that fictitious commodities (and AI and data could qualify as such) dis-embed from the social, the more likely it will be that counter-movements will emerge that attempt to return these commodities to social relevance.[15] Any socio-political enterprise in which AI is employed to ensure equality and counter discrimination will need to confront the exclusionist economic pressures that marketize technology, and the data they produce in neo-liberal individualist wealth creation imperatives.[16]

> *"As our work on COVID surveillance establishes, there is little doubt that community apprehension about AI, in large part as suggested above, revolved around arbitrary power. To overcome this concern, standardisers and the engineering community look for endorsements of safety and robustness through mechanical risk minimization."*

During the pandemic, many nations employed AI-assisted surveillance technologies, and mass data sharing as key components to their health/safety control strategies. In several of these societies, the structural discrimination directed, for instance, against migrant worker populations made these communities of high risk.[17] Overcrowded living conditions, poor hygiene and diet, limited access to sufficient preventive health-care services, and a general myopic social exclusion meant that when the risks were recognized too late, harsh surveillance regimes, mass lockdowns and incubation, accompanied by significant psychological stress from the uncertainty of future family contact, harshly evidenced that these vulnerable groups were further discriminated against with AI-assisted technology as part of the state's armory.[18]

[14] Mark Findlay, (2017), chap. 5.

[15] Karl Polanyi, *The Great Transformation: The political and economic origins of our time,* (Boston: Beacon Press., 2001)

[16] Mark Findlay, *Globalisation, Populism, Pandemics and the Law* (Cheltenham: Edward Elgar, 2021).

[17] Jane Loo, Josephine Seah and Mark Findlay, "The Vulnerability Project: Migrant Workers in Singapore," *CAIDG Research Paper*, (29 January 2021), https://papers.ssrn.com/abstract=3770485

[18] Seah Loo and Findlay, (2021).

Many of the confinement requirements could not have been achieved without AI-assisted monitoring and body tracing.

Some more authoritarian states engaged in mass data sharing policies where private sector commercial databases and state repositories surveilled populations to an extent previously unimagined.[19] While this process may have appeared universal in its coverage, it had a discriminatory impact on workers traveling across provincial borders, and those not able to enjoy the protective benefits of working from a domestic setting. In addition, due to the sometimes faulty analysis of this merged data, not insignificant numbers of citizens were excluded from travel and employment through false readings.[20]

More worrying, as with vaccine availability worldwide, the selective access to AI-assisted control technologies due to poverty, displacement, and technological marginalization meant that state responses particularly to transient and racially victimized populations were often brutal, delayed, and counterproductive.[21]

Discrimination was also visited on the elderly[22] who due either to their limited digital literacy, possession of dated technology, or institutional confinement where family contact was reduced to on-line encounters, fell outside the beneficial influence of AI-assisted surveillance instead it enabled freer association and movement for the wider population if technological compliance was available and ensured.

It would be incorrect to lay blame for all the discriminatory consequences of COVID control and the disquiet these produced[23] on AI tech and big data. However, these instances summarized above are only some of the instances where the exercise of state/private sector surveillance powers, reliant on AI and its data, challenged the legitimacy of authority styles.[24] Few are the examples where communities were

[19] Cate Cadell, "China's Coronavirus Campaign Offers Glimpse into Surveillance System," *Reuters*, 26 May 2020, https://www.reuters.com/article/us-health-coronavirus-china-surveillance/chinas-coronavirus-campaign-offers-glimpse-into-surveillance-system-idUSKBN2320LZ

[20] Helen Davidson, "China's Coronavirus Health Code Apps Raise Concerns over Privacy," *The Guardian*, 1 April 2020, https://www.theguardian.com/world/2020/apr/01/chinas-coronavirus-health-code-apps-raise-concerns-over-privacy

[21] Julie E. Cohen, Woodrow Hartzog and Laura Moy, "The dangers of tech-driven solutions to COVID-19," *Tech Stream,* 17 June 2020, https://www.brookings.edu/techstream/the-dangers-of-tech-driven-solutions-to-covid-19; Susan Landau, Christy E. Lopez and Laura Moy, "The Importance of Equity in Contact Tracing," *Lawfare,* 1 May 2020, https://www.lawfareblog.com/importance-equity-contact-tracing

[22] Tessa Oh, "Covid-19: NGOs Step up Outreach Efforts, as Seniors Try to Stay Occupied While Isolated at Home," *Today*, 18 October 2021, https://www.todayonline.com/singapore/covid-19-isolated-seniors-tell-how-they-stay-occupied-while-staying-home-ngos-step; Akshita Nanda, "Seniors Stuck at Home, Caught between Loneliness and Fear of Covid-19," *The Straits Times*, 9 October 2021, https://www.duke-nus.edu.sg/care/news-events/news/articles/articles/seniors-stuck-at-home-caught-between-loneliness-and-fear-of-covid-19-(straits-times-premium)

[23] We investigated community disquiet and distrust in the context of these technologies and measures. Six broad themes were identified in total. See: Alicia Wee and Mark Findlay, "AI and Data Use: Surveillance Technology and Community Disquiet in the Age of COVID-19," *CAIDG Research Paper*, (26 October 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3715993

[24] Jane Loo and Mark Findlay, "Rule of law, legitimacy and effective covid 19 related technologies: examining state

engaged in the deployment of AI surveillance and mass data sharing.[25] In contexts where social capital tolerated these intrusions for the sake of utility there is still evidence that trust in the institutions and processes of governance was damaged.[26]

If equality in deployment and application of AI tech and the use of its data becomes ranked as an important principle influencing the machine/human interface, then this will be best achieved through the creation and maintenance of relationships of trust in communities.[27] Additionally, via the operation of these relationships trust can become a regulatory force ensuring the governance of AI by recognizing accountability to its recipients.

RoL, Inclusion and COVID Surveillance and Control Policy - Vulnerability and the Exacerbation of Discrimination

Rule of law (RoL) is recognized worldwide as a fundamental democratic discourse[28], and compliance with its principles is a claim for democratic legitimacy.[29] As such, the application of AI in both economic and political contexts (if that application is to be deemed to have a democratizing potential) can be fairly contrasted against rule of law language, if one searches for a relationship between AI and democratic values. Rule of law represents both normative and applied realms of democratic governance with its many principles have been interpreted from literalist/substantivist, to thick and encompassing impacts, and this spectrum of meaning adds to its political if not its instrumentalist utility.[30] There are also significant correlations between RoL principles and the ethics espoused as universally ascribed foundations for AI governance.[31]

---

authority, power, legitimacy and citizens trust during covid 19," (forthcoming).

[25] The same cannot be said of recent and significant instances where community participation has been active in environmental protection spheres such as power conservation and waste management.

[26] Particularly in those instances where utility is undermined. See for example: Paul Mozur, Raymond Zhong and Aaron Krolik, "In Coronavirus Fight, China Gives Citizens a Color Code, With Red Flags," *The New York Times,* 7 August 2020, https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html; Helen Davidson, (2020).

[27] Li Min Ong and Mark Findlay, "A Realist's Account of AI for SDGs: Power, Inequality and AI in Community," in *Artificial Intelligence (AI) for the Sustainable Development Goals (SDGs): Socially Good AI, Artificial Intelligence for the Sustainable Development Goals (Tentative),* ed. L. Floridi & F. Mazzi. (Springer Nature (Tentative)). Also see, https://binghamcentre.biicl.org/projects/the-role-of-good-governance-and-the-rule-of-law-in-building-public-trust-in-data-driven-responses-to-public-health-emergencies

[28] Sandra F. Joireman, "Colonization and the Rule of Law: Comparing the effectiveness of common law and civil law countries," *Political Science Faculty Publications*, Vol. 64 (2004), https://scholarship.richmond.edu/cgi/viewcontent.cgi?article=1077&context=polisci-faculty-publications. See also: Jeremy Waldron, "The Rule of Law," *Stanford Encyclopedia of Philosophy*, (22 June 2016), https://plato.stanford.edu/entries/rule-of-law/

[29] Martin Krygier, "The Rule of Law and State Legitimacy," accessed on 29 March 2021, https://oxford.universitypressscholarship.com/view/10.1093/oso/9780198825265.001.0001/oso-9780198825265-chapter-7

[30] Danielle Watson, Ariel Yap, Nathan W. Pino and Jarrett Blaustein, "Problematizing the Rule of Law Agenda in the SDG Context," *Crimrxi,* (8 December 2020), https://doi.org/10.21428/cb6ab371.ba393049

[31] Ellis Paterson and Gemma McNeil-Walsh, "Catching up with the Debate: Artificial Intelligence & the Rule of Law," *Reconnect*, 14 October 2019, https://reconnect-europe.eu/blog/aiandrol-patersonmcneilwalsh/

In institutional and process forms, the RoL functions as a mechanism of constraint[32] thereby ensuring that governance authority and powers exercised are legitimate. Legitimate democratic powers are endorsed by citizens and such endorsement is likely to be enhanced through a trust in authority that subscribes to rule of law.[33] As with the reservations expressed above about limiting governance considerations to state or market entities and functions, rule of law is at least a two-way dynamic. Whether its legitimating capacity is claimed through governance that is state-initiated, MNC-sponsored (as through ethics or corporate social responsibility), or is communitarian in its origins and operative responsibilities, the citizen is a critical endorsing agent. Further, rule of law as a trust generator is majoritarian – meaning that most citizens (and their communities) need to have trust in its functioning, or at least those in the community most affected by the exercise of the power in question. As such, while rule of law may sit comfortably with state-centered liberal democratic models of governance and their reliance on one-person-one-vote, its influence is not confined to that governance setting. If RoL is to have a governing impact over AI, then its application to private power, problematic as this may seem, needs to be extrapolated.

AI in the community that adheres to the principles of the rule of law will satisfy democratic values and norms, and operational communitarian governance will be seen (at the very least) to possess democratic/representative legitimacy provided that citizens (individually and collectively) trust AI's purposes and priorities through deployment.

Predominantly, communities are concerned by the dangers and risks AI poses within their midst, rather than embracing its benefits.[34] This reality, in part, is based on the misguided notion (triggered by science fiction images) of machines with moral agency and autonomous consciousness acting in an arbitrary fashion. Rule of law compliance (by AI in the exercise of power) tempers and curbs the arbitrary exercises of technology powers, which in turn enhances citizen trust, linking back to the importance of initiating trust within communities between human recipients and AI.[35]

---

[32] In tempering both state and non-state arbitrariness

[33] Powers that are legitimate are more likely to be endorsed by citizens leading on to greater compliance with State control measures, see: Jane Loo and Mark Findlay, "Rule of law, legitimacy and effective covid 19 related technologies: examining state authority, power, legitimacy and citizens trust during covid 19," (forthcoming).

[34] Michael Zimmer et al., "Public Opinion Research on Artificial Intelligence in Public Health Responses: Results of Focus Groups with Four Communities," *AAAS Center for Public Engagement with Science and Technology*, (10 August 2021), https://www.aaas.org/sites/default/files/2021-09/AI%20in%20Public%20Health%20Focus%20Groups%20-%20Final%20Report%20with%20Appendix.pdf;
Kelsey Piper, "The American public is already worried about AI catastrophe," *Vox,* 9 January 2019, https://www.vox.com/future-perfect/2019/1/9/18174081/fhi-govai-ai-safety-american-public-worried-ai-catastrophe

[35] Wong Zhang and Findlay, "Trust and Robotics: A Human-centric Pathway to Robots in Community," (forthcoming).

Communities are both inclusive and exclusionist in the forms of shared governance under which they operate. In Cotterrell's reading of community, its existence is determined by bonds of trust[36], and the assumption therefore, is that for AI to be an active partner in community it must participate in, and maintain trusted relationships. As with the question of where AI is, and is not, appropriately located in communities, there will always be occasions either where the human recipient will deny trust in AI technology or that technology will be unable contextually to create or continue trusted relationships. On inclusion – RoL actionability (in execution and institutional grounding terms) lends to overall assurance of positive citizen engagement by offering some enforcement options when abuse of power requires remedies.[37]

Rule of law was said to be transported along with European colonization of the South World and the benefits for good governance were collateral.[38] Unfortunately, Westminster-style parliamentary democracy has not always taken root in positive governance environments, which have then tended to advantage traditional elites, foster institutional corruption, and enable North World trade exploitation.[39] Democratic normative frames have also suffered transition into vulnerable economies still impacted by the ravages of autocratic neo-colonial rule. If AI deployment accompanies relationships of technological dependency within vulnerable economies then rather than providing a democratic platform, AI will rather duplicate neo-colonial economic dominion.

Regarding the democratic value of equality – the principles of equality and non-discrimination that are foundational to the rule of law ensure that people are treated with dignity and respect. It would be logical to assume that the agents offering or denying dignity and respect, and therefore using data in a discriminatory fashion, could be AI-assisted technologies. Returning to COVID control policy, in many styles of governance authority, particularly in the South World where rule of law buffers are less robust, traditional discrimination based on race, gender wealth and power formed the backdrop for inequitable intrusions into citizen liberties.[40]

Unfortunately, rule of law's rejection of arbitrary and abusive powers had

[36] Roger Cotterrell, "Law, Emotion and Affective Community," *Queen Mary School of Law Legal Studies Research Paper*, (15 July 2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3212860

[37] Mark Findlay, "Ethics, Rule of Law and Pandemic Responses," *CAIDG Research Paper,* (30 July 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3661180

[38] Mark Findlay, (2017), Chapter 5.

[39] Mark Findlay and Si Wei Lim, *Regulatory Worlds: Cultural and Social Perspectives when North meets South* (Cheltenham: Edward Elgar, 2014).

[40] Shakir Mohamed, Marie-Therese Png and William Isaac, "Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence," *Philosophy & Technology*, Vol. 38 (July 2020): p. 659–684, https://doi.org/10.1007/s13347-020-00405-8

limited impact in many instances of COVID control intrusions, if the measure of legitimate authority was more about the utility of the powers' applications than more democratic normative concerns.[41] Opposition to arbitrariness requires aligning the exercise of powers (and facilitating technologies) in accordance with RoL principles to prevent its uncontrolled, unpredictable, and disrespectful exercise. In turn, this will enhance values of inclusion and equality that are fundamental for democratic governance. On the other hand, technologies and powers that are illegitimately exercised, (tending to be disrespectful or neglectful of citizen inclusion, equality, and vulnerability) or excessively or abusively exceeded, will be received poorly and this will have a negative consequence on both the technology proposed and the legitimacy of any dominant authority form. Obviously, this reaction depends on the manner in which social capital has an independent voice.

The rule of law safeguards the value of inclusion because of the emphasis it places on the principle of legal certainty. Legal certainty is about eliminating the application of arbitrary unpredictable powers that make laws/measures/decisions taken inaccessible, intelligible, and unclear to those individuals affected. Compliance with the principle of legal certainty prevents the arbitrary exercise of unpredictable powers because it elevates subjects in a better position to know, predict, understand, comply and most importantly, *participate meaningfully* in decisions involving them. Occasions of indecisive authority, lack of coordination, and backflipping of decisions in today's pandemic context are examples of how the State (and other authority agencies) act in an unpredictable and legally uncertain manner. In these cases, citizens are excluded from participating since they are unclear on how to act and cannot organize their affairs in accordance with the relevant rules in force. This inability to participate meaningfully will impact their willingness to comply not simply because citizens feel excluded but also because their perception of legitimacy and trust in the expertise of government actions (and its control technology) falters.

Meaningful inclusion[42] is also about accountability and transparency in the sense that when powers are exercised arbitrarily, there are avenues and means made available for subjects to challenge actions that interfere with their rights/interests. The emphasis on access to justice in RoL parlance encourages transparency and accountability from those who would exercise power through technology. This leads to the early identification and remedying of inconsistent laws and abuses in their name.

[41] Jane Loo and Mark Findlay, "Rule of law, legitimacy and effective covid 19 related technologies: examining state authority, power, legitimacy and citizens trust during covid 19," (forthcoming).

[42] Discussed in the context of inclusive self-regulation dependent on informed participation see Mark Findlay and Josephine Seah, "Data Imperialism: Disrupting Secondary Data in Platform Economies Through Participatory Regulation," (29 May 2020). *SMU Centre for AI & Data Governance Research Paper No. 2020/06*, Available at SSRN: https://ssrn.com/abstract=3613562 or http://dx.doi.org/10.2139/ssrn.3613562

AI in Community – Governance and Legitimacy

In earlier work,[43] we proposed that *AI in community* can function as a governance framework for AI systems, but have yet to sufficiently develop the stages in its causality. So, what follows can be treated as thoughts in progress. Community-mindful AI design and deployment can initiate relationships of trust within community recipients, but more importantly, democratic values of inclusion and equality, which underpin our idea of AI in community, will provide the legitimacy platform critical for citizen endorsement in modern democratic (equitable/inclusive) governance and its technologies. In our work on disquiet arising from the exercise of AI-assisted surveillance in pandemic control, we initially proposed an absence of citizen inclusion in the policy formulation and application as interfering with trust and efficacy.[44] That view was later refined to suggest that an absence of informed and participatory inclusion in various stages of policy development and deployment would create distrust in other elements of authority required for the legitimacy of the powers exercised.[45]

At least in those styles of authority where citizen/community trust is a necessary or even powerful legitimator,[46] RoL provides a useful frame of analysis for the democratic governance of emerging technologies since close adherence to its values will enable and initiate citizens' trust. The initiation and maintenance of such trust foundations is important in societies seeking a model of AI governance that is inclusive and located *within* and *for* the community. More importantly, adherence to its principles will also operate to promote the legitimacy of the technology, which in turn legitimizes the democratic powers and authority of the State (and other relevant governance authorities). The converse is also true – legitimate authority and powers will legitimize and endorse the technology in use.

Moving from the undeniable relationship between community trust and legitimate authority is the more pinpoint focus of governing the exercise of often intrusive AI-assisted powers. As our work on COVID surveillance establishes, there is little doubt that community apprehension about AI, in large part as suggested above, revolved around arbitrary power. To overcome this concern, standardisers and the engineering community look for endorsements of safety and robustness through

[43] Findlay and Wong, "Trust and Regulation".

[44] Wee and Findlay, "AI and Data Use: Surveillance Technology and Community Disquiet in the Age of COVID-19".

[45] Jane Loo and Mark Findlay, "Rule of law, legitimacy and effective covid 19 related technologies: examining state authority, power, legitimacy and citizens trust during covid 19," (forthcoming); Findlay and Wong, "Trust and Regulation,".

[46] Jane Loo and Mark Findlay, "Rule of law, legitimacy and effective covid 19 related technologies: examining state authority, power, legitimacy and citizens trust during covid 19," (forthcoming).

mechanical risk minimization.[47] Extending this search for satisfying risk into the community, and working with the importance of trust between AI and human recipients, requires an a priori appreciation that compliance with standards or safety ratings may not counter adverse or irrational popular wisdom.

However, like ethics as an AI governance frame, the democratic values enunciated in rule of law and the SDGs, requiring certain forms of compliance will not be satisfactory if these are left in the self-regulatory domain.[48] Recipient citizens and their communities have a vital role to play not only in trusting AI, but in first knowing, understanding and making accountable the achievement of principles and values on which trust is based. Communities can provide the 'quality control' referent for AI governance, and fortuitously AI technology and responsible access can facilitate this responsibility. Time does not allow for a developed justification of this assertion, but in the same way AI can deny, or bias access to information through algorithmic exclusion, algorithms can red-flag biased interpretations, open up data access through information looping and provide broad and basic frameworks for citizen/community engagement with the knowledge on which their identities rest.[49] In such ways, AI and data usage, if embedded in empathetic community relationships, can offer mechanisms and fail-safes that keep the authors of authority and the exercisers of power, true to their democratic discourse.

By recognizing individual vulnerability and eliminating discrimination, the democratic values of equality and inclusion are more likely to be realized. For instance, equality is consistently highlighted as a core element of the RoL, and it is directly relevant and even transplantable in this context. Equality is about ensuring that authority governance modes and the power they exercise (in public and private spheres through technology) must not only refrain from discriminating against certain groups/individuals but must also seek to protect these groups against discrimination based on certain protected characteristics. Equality before the law should not be proclaimed when, as the context of this pandemic and its increasingly stringent control measures, authorities are given preferential treatment by virtue of their domination of technology and data access to disregard equality under conventional law.[50]

---

[47] Zhang, Wong and Findlay, "Trust and Robotics: A Human-centric Pathway to Robots in Community," (forthcoming); Megan Nichols, "Safety Is a Priority When Designing Collaborative Robots," *Robotshop,* 20 March 2019, https://www.robotshop.com/community/blog/show/safety-is-a-priority-when-designing-collaborative-robots

[48] See Findlay & Seah, "An Ecosystem Approach to Ethical AI and Data Use: Experimental reflections," *CAIDG Research Paper*, (13 May 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3597912

[49] Mark Findlay and Li Min Ong, "Reflection on Wise Cities and AI in Community: Sustainable Life Spaces and Kampung Storytelling," *ASEAN Law Research Network e-Paper Series*, Vol. 1 (2022) (Forthcoming), (Sustainable Development and Commerce in ASEAN Cities Conference (12 Nov 2021)).

[50] This is not to suggest that I believe laws are equal in their creation, intentions or applications, beyond normative exhortation – see Mark Findlay, (2017), Chapter 5.

Where the discussion of AI in community is presently at its weakest, impacting on any supposition about AI being made complicit in ensuring good communitarian governance, then there remains the need:

- to develop a more complex, layered and iindividualized understanding of the AI/human relationship and what makes it trusted
- to better appreciate the factors which incentivize or de-incentivize communities to embrace AI beyond shared self-interest
- to materialize the forces that will influence AI sponsors/designers and customers so as to accept as a significant driver for AI deployment community benefit and prioritization outside the market contest and,
- to position AI so that it is more than some patient/impotent recipient of community freewill and to entertain a more proactive (if not autonomous) role for AI in social bonding.

However, even at the current stage of thinking, the causal chain for AI and democratic values might look something like this:

AI in community (based on relationships of trust) to enunciation/compliance with democratic values at design/deployment to recognition in community that AI/data use represents these values. Then, AI/data use is legitimated the AI/data use provides the technology for communities to audit the use by governance/authority of technology/data power community thereby becomes the 'governor' of democratic values in AI/data use and their application through governance/authority.

Conclusion

We should be mindful, when looking into how AI and democratic values can or should intersect, not to reduce the analysis to a tick-box exercise or some checklist of generalized and abstract principles. Through identifying equality within the SDGs and inclusivity in rule of law as indicia of the democratic values of equality and inclusivity, it is essential to see the governance of AI (and AI's potential to advance human-centric, communitarian governance best practice across its ecosystem) as a *teleological* endeavour. For instance, any interest in RoL manifest in AI design and deployment imperatives requires a consideration of function as well as value. How can rule of law curb the arbitrary exercises of power via AI technology and data usage, and how can advancing the SDGs offer more sustainable communities as foundations for governance against arbitrary or abusive uses of technology and data?

The case study of AI surveillance discrimination against vulnerable groups gives a snapshot of how arbitrariness is often linked to inequality and an exclusion and the exploitation of powerlessness. That said, meaningful guidance can still be found in an *anatomical* account of the rule of law and the sustainable development goal aspirations. At the very least, normative confirmation provides some action-oriented direction for how arbitrariness can be tampered – including what sort of formal/procedural characteristics need to be safeguarded and what communitarian governance frames should be put in place for that purpose. Conceding that the arbitrary exercise of autonomous power is high on the list of why communities do not trust AI (and its authority sponsors), it is a practical imperative to include AI in the achievement of democratic values that run counter to arbitrary power, and in which AI/data usage can be positive governance moderators.

ARTIFICIAL INTELLIGENCE, SOCIETY AND DEMOCRACY

Artificial Intelligence (AI) is a popular discipline within Computer Science that has direct implications in society. Recent advancements in technology, ubiquity of computing and explosion of data in various applications have made it significantly cheaper to develop AI, infuse it in various dimensions of daily life, and digitally transform industries. While almost all disciplines of technology have their own merits and pitfalls, AI poses an important challenge to how our society is formed. Previous technological advancements like the first waves of the industrial revolution made it easier to produce goods and services while reducing the cost, expanding the value of these revolutions to masses. AI, on the other hand, makes it easier for computers to act cognitively, taking over some of the most "human" aspects of our lives, which necessitates a new Social Contract among individuals, corporations, nation states, international organizations, based on the concepts of Responsible AI, Trust and Upskilling. This text builds on a brief definition of Artificial Intelligence (AI), shares some examples of AI use cases as well as limitations, provides recommendations on creating a relationship between democratic values and Responsible AI principles, and suggests how to support these developments through national and international governance mechanisms, including upskilling of citizens.

Cavit E. Yantaç*

TPQ
Winter 2021/22

* Cavit E. Yantaç is Deputy GM, Customer Success Leader at Microsoft Turkey and holds visiting instructor positions at Bogazici University Executive MBA program and Bilgi University IT Law Institute.

Artificial Intelligence (AI) is a popular, widely analyzed, and speculated area of Computer Science that has direct implications in society, through various legal, regulatory, economic, and educational dimensions. With recent developments in the world of technology (including cloud computing, explosion of big data, prolific enhancements in AI algorithms), the cost of developing and distributing AI applications has diminished significantly, democratizing the once expensive and challenging realm of AI throughout the world. As the impact of AI is now visible across different disciplines (examples include law, media, economics, defense, etc.), it is crucial to evaluate, and proactively plan for, a world where AI coexists with human organizations.

Brief Definition of Artificial Intelligence (AI)

What is intelligence and what makes humans intelligent? Can non-living beings demonstrate intelligence; or in other words, is it possible to have "artificial" intelligence, rather than "natural" intelligence"? If the answer was positive, what would a good example of artificial intelligence be? The first philosopher to contemplate a non-living intelligent agent that could answer questions directed to itself was Descartes in 1600s.[1] And while the first computers were designed in the 19th century (the first mechanical computer was created by Charles Babbage in 1822), advancements in machinery and electronics later led to the development of more powerful computers after WW2, and microprocessors and PCs after 1960s and 70s. As the size and cost of computers decreased, their processing capabilities increased constantly and exponentially. Even more important was the development of Internet, which enabled these devices to connect to each other, scaling their capabilities immensely in 1990s and the 21st century.

Along with these developments, computers have always depended on the concept of an IPO model (Input, Processing, and Output): receiving inputs (e.g., touching a screen, pressing a keyboard button), processing them (e.g., making a calculation, formatting text on an email) and providing outputs (e.g. displaying an image on a screen, writing the calculated number on a disk). How a set of inputs would be "processed" would depend on an algorithm, exactly defined, and explicitly codified by a "programmer". Example: If the user touches a particular button on the keyboard, do X; if the user touches another button, do Y, and so forth. Interestingly, this IPO model is very similar to how our brains process information. The nervous system of living organisms relies on inputs (e.g. sights, sounds, etc.), processing capabilities (e.g. reading and understanding text) and outputs (memorizing the text or speaking it out), much the same way as how we've developed computers. This

[1] Minsoo Kang, The *Mechanical Daughter of Rene Descartes: the Origin and History of an Intellectual Fable* (Modern Intellectual History, 2017).

similarity between living organisms and computers has led to the question of whether computers can act similarly to other intelligent agents such as humans, dolphins, ants, etc., which bears even a further question of what intelligence is. These questions have been answered by various computer scientists, philosophers and neuro-scientists since 1950s.

"Due to an explosion of data that resides throughout the internet (including massive data centers that power the cloud as well as smart edge devices), enhancements in graphical processors, the willingness of researchers throughout the world to share their learning algorithms through open-source projects have all led to further enhancements in Machine Learning."

Max Tegmark, in his book "Life 3.0 Being Human in the Age of Artificial Intelligence", defines intelligence as "the ability to accomplish complex tasks" and breaks this down to three main components: processing, memory and learning.[2] From both a processing power and memory perspectives, even though humans have been one of the most powerful living beings, they have recently been outpowered by computers due to advancements in microprocessors. "Learning", as a concept, is challenging for computers, though – as noted above, computers have relied on very explicitly written definitions of exactly how they should behave in each instance of running a program. However, because of the complexity of our universe, this exact and explicit definition of all combinations of all objects is very challenging, if not impossible, necessitating "learning" capabilities. Machine Learning (ML), therefore, is a key construct that enables artificial intelligence. Due to an explosion of data that resides throughout the internet (including massive data centers that power the cloud as well as smart edge devices), enhancements in graphical processors, the willingness of researchers throughout the world to share their learning algorithms through open-source projects have all led to further enhancements in Machine Learning. Examples of such learning capabilities include recent announcements on computers reaching human parity in image recognition[3] or speech recognition.[4]

[2] Max Tegmark, *Life 3.0 Being Human in the Age of Artificial Intelligence* (Knopf, 2017).

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, "Deep Residual Learning for Image Recognition," (2015), https://arxiv.org/abs/1512.03385

[4] W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, D. Yu and G. Zweig, "Achieving Human Parity in Conversational Speech Recognition", (2016), https://arxiv.org/abs/1610.05256

Benefits and Pitfalls of AI

The above-mentioned developments show that it is now possible for computers to interact with humans at a level almost indistinguishable from other humans in some dimensions of cognition, which has created a lot of interest in possible uses of AI. A few examples of positive use cases of AI are given below:

- **Image Processing, Detection and Classification:** As examples, AI can be used to process streams of videos to classify objects to understand if an autonomous car is faced with a traffic light, or a person, or a cat; or recognize instances that pose threat for employee safety where an employee is not wearing a helmet in a factory. Similarly, AI-based systems can be used to analyze MRI scans of millions of cancer patients, and over time, detect patterns to support radiologists' ability to detect cancer early.
- **Natural Language Processing:** AI can be utilized in call centers to understand customers' complaints or swiftly browse through huge volumes of text generated in judicial systems to gather information for lawyers.
- **Forecasting and Optimization:** AI can also be used by large retailers or financial institutions to forecast demand for a particular product and guide their customers to optimal products and services. It can be used to predict if a particular customer of a bank is likely to default, guiding the bank and its customer to new products before such a default happens.

Almost all industries, from education to agriculture, pharmaceuticals, or financial services, are prone to using AI in various forms. It would be naïve, though, to assume that it's all a glorified road to using AI in everyday life. What are potential pitfalls of using AI in an effective manner in all dimensions of life? Just a few points of caution are summarized below:

- **Data Quality and Quantity:** AI depends heavily on the quality and quantity of data submitted to it, and if left unattended, data in its natural form generally includes significant amounts of quality issues, including biases. To account for these issues, a successful AI implementation includes multiple steps of data cleansing, which typically relies on data scientists who are also succinct subject matter experts in the field they're analyzing.
- **AI Interpretability:** Contrary to a simple application that uses a set of explicit calculations, the above examples of image processing or natural language processing generally use deep learning algorithms that rely on

multiple layers of complex networks of data, which makes it difficult for developers to understand exactly how those algorithms work.

- **Privacy:** Due to the amount of data collected for effective AI implementations, it becomes easy for developers of AI to have direct oversight on individuals' behaviors or characteristics. A good example of privacy considerations is observed when governments or corporations install cameras in the street to detect traffic patterns, but do not take the relevant steps to mask-out or anonymize faces or number plates.

"As listed above, the main pillars of Responsible AI would include at least fairness, inclusiveness, reliability, safety, transparency, privacy, security, and accountability."

These developments all bear very important questions: As AI-powered systems support (or potentially replace) humans in critical decision-making circles, such as Courts, Security or Healthcare, how can we develop strong, ethical, and accountable mechanisms that protect the privacy of citizens? How can we secure international cooperation on the use and transfer of data, AI algorithms, and their use in decision-making across borders? How would a society that upholds democracy use AI? What kinds of skills need to be developed for effective use of AI within society?

Democratic Values and Responsible AI

Democracy, which can be defined as the peoples' authority to decide on legislation directly, or elect representatives to decide on such legislation, depends on critical values such as liberty, equality, justice, freedom of speech, voting rights, etc. These values, upheld by individuals or groups of individuals, and protected through explicit codes of laws and regulations, form the basis of decision making at both micro and macro levels, and facilitate inclusive decision making. And if a certain judicial or economic decision, for instance, were to be made by systems empowered by AI, how could we assure individuals that their preferences are included in decision making, while securing their rights to privacy and security?

The answer is likely to arise from a new Social Contract, to be devised among sovereign states, non-governmental organizations, corporations, academia, interest groups and individuals. This social contract would need to account for at least the following principles on the use of AI, which may be summarized as the "principles of Responsible AI"[5]:

[5] For further reading on Responsible AI, see "Responsible Use of Technology: The Microsoft Case Study", (2021).

Fairness and Inclusiveness

Fairness and inclusiveness are critical tenets of core democratic values such as equality, justice, and freedom of speech. We should expect AI systems to be based on the principle of treating all people fairly, ultimately improving the overall fairness in any judicial, political, legal system today. AI systems can allocate or withhold opportunities or resources based on input provided to them – presenting a possibility of over- or under-representation of certain individuals or groups. An example of under-representation of a social construct is where disadvantaged rural communities have fewer opportunities to access systems using AI, reducing their impact on the decisions driven by such systems. An AI-based decision-making cycle should ensure that the voices of everyone are heard, and that their contributions are provided to the system as relevant inputs.

Reliability and Safety

Democratic sovereign entities aim to provide physical, psychological, and mental safety for all citizens. With that main goal in sight, AI systems should perform reliably and safely; and should be open to constant improvement based on human feedback. An AI project that provides a successful result in one city, country, or social group may not necessarily provide the same level of efficacy in another city, country, or social group. Developers of AI should incorporate a feedback mechanism for humans to constantly update the models used in AI. Similarly, AI applications should be developed with mandates to cease to operate when faced with situations that may result in any harm to humans.

Transparency

Transparency in AI is critical because it enables users of AI to trust the system, and trust is at the core of any democratic, social construct. As stated above among pitfalls of AI, though, AI-interpretability may be hard to achieve due to the technical complexity of algorithms used. Companies who develop AI applications should make concerted efforts in documenting and proactively sharing their methods, opening their code to external communities, and being open about the limitations of their systems.

Accountability

Despite all the complexity of AI applications that can be hard to interpret, humans should be held accountable for the results of AI systems. Accountability also includes structures that are in place to ensure the developers are acting in accordance with the core principles of ethics, rule of law, and democracy. Developers

of AI should have clear principles on how they develop, sell, and advocate for their applications; and sovereign entities should have clear guidelines in protecting their citizens against pitfalls.

Privacy and Security

Privacy is a fundamental right that protects citizens against any power that may have oversight on their preferences, actions, or statements. As AI and Machine Learning all rely on using data to model the universe and learn, and that increasing reliance on data adds a new level of complexity, and new requirements for keeping it secure. Developers of AI should have clear boundaries on what they can access as they develop an application, train a model, or use it in real life. As an example, an AI developer who's working on an application that detects credit card fraud, for example, should not have access to all personally identifiable information of the user, and should only have access to a "masked" set of data. Data origin and lineage, use of data internally or externally, or data corruption considerations should be constantly evaluated by developers and users of AI.

A successful example of international efforts to secure privacy is the General Data Protection Regulation (EU) 2016/679 (GDPR), a regulation in the EU and European Economic Area (EEA) on data protection and privacy. Adopted in 2016, and enforced since 2018, it is a directly binding and applicable regulation that protects the privacy of all citizens of EU and EEA and applies to all enterprises regardless of their origin and data subjects' citizenship. GDPR has been applied as a role model for many other privacy-protection regulations across the world, including Brazil, Chile, Japan, South Korea, South Africa, Malta, and Turkey. It is based on the idea that personal data may not be processed unless there is at least one legal basis to do so, such as a person's consent to such processing, or a data controller's legal obligations.[6]

More recently, the European Commission released a proposal for the regulation of not just Data, but also AI[7], which aims to "build an ecosystem of excellence in AI and strengthening the EU's ability to compete globally" while ensuring that "Europeans can trust the AI they are using." This effort shows that a single nation state or a single governing body is not sufficient to build an environment of trust, and that international cooperation is needed to secure especially privacy, in the age of social media, ubiquitous computing (including smart phones, smart watches, smart buildings, smart cars, etc.).

[6] For further reading on "General Data Protection Regulation", see https://eur-lex.europa.eu/eli/reg/2016/679/oj

[7] For the European Commission's Proposal Regulation on AI, see https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

As listed above, the main pillars of Responsible AI would include at least fairness, inclusiveness, reliability, safety, transparency, privacy, security, and accountability. For a society to effectively stand on these pillars, all relevant actors that develop and govern AI applications should jointly agree on a contract that endorses these pillars, which, as stated above, can be constructed through a social contract. The government's role here would be to protect its citizens while maintaining an environment where innovation is possible. The private sector's role would be to develop applications with clear governance mechanisms to oversee how AI applications adhere to the principles mentioned here. As potential users or developers of AI, government entities should also apply at least the same principles as the private sector, and at least with the same level of openness and trust.

Building an International Environment of Trust in AI

Building an environment of trust in AI requires a social contract, as stated above. However, compared to previous technologies which could be monitored on a national level (such as the TV or telecommunications industries), this new social contract needs to expand through national boundaries as no one state, or country, or company controls the AI ecosystem, which interweaves a convoluted web of hyperscale cloud operators, large scale developer companies, multinational service providers, regional or local telecommunications providers, local or state governments or supranational entities like the EU or NATO.

A widely discussed question of legal responsibility in AI and robotics lays the problem of international collaboration clearly: if an autonomous car has an accident while in a self-driving mode, who would be accountable for penalties that will be incurred (an autonomous car is essentially a robot with advanced AI capabilities)? Part of the answer lies in the Vienna Convention on Road Traffic, an international treaty designed to facilitate international road traffic, which was signed in 1968 and entered into force in 1977.[8] The Vienna Convention acts as the foundation for establishing standard traffic rules among contracting parties, including the fundamental principle that a driver is always fully responsible for the behavior of a vehicle in traffic. If a car is in a self-driving mode, can the driver in one country be held accountable for the decisions of an AI application that could have been developed in another country, and that resides on a chip that was produced in another country? Like the Vienna Convention, nation states have recently come together to discuss potential amendments to the original text of the Vienna Convention.

Similar international conventions and agreements are needed to control how AI

[8] For Vienna Convention and the standards on international traffic rules, see https://treaties.un.org/doc/Treaties/1977/05/19770524%2000-13%20AM/Ch_XI_B_19.pdf

systems are developed, deployed, and updated, across different industries and domains, such as medical sciences and public health, law and judicial bodies, mainstream media and social media. Such international agreements need to recognize that corporations that now build AI applications seldom work in a single-country environment, and therefore should protect the rights of citizens of each country within a global context. At the same time, these international agreements should also focus on fostering innovation, rather than stifling all AI applications.

Educating a Society for AI

On one hand, AI poses a threat to uneducated or unqualified masses due to the power of automation it brings on many industries. On the other hand, though, it also provides immense new opportunities for maximizing economic output and improving the lives of everyone. To effectively utilize the benefits of AI while minimizing the potential negative impacts, societies need to take deliberate steps to educate their citizens so they can coexist with AI-based services.

The first stream of thought, when thinking about such education, would include the upskilling of citizens in the areas of computer science or data science, for instance, as potential developers of AI. However, the cost of such education and its prerequisites (such as a strong foundation in mathematics or statistics or the availability of computing for everyone) makes it impractical for most nation states to deploy nationwide for all students and adults.

Therefore, a good second choice would be to educate all citizens in various fundamental skills (but mostly on the social side), which cannot easily be overtaken by computers and AI systems. AI can provide individuals and organizations with huge new cognitive capabilities, but humans (at least for the foreseeable future) have evolved over hundreds of thousands of years to become very effective in social skills and have the upper hand in intuition, abstract and conceptual thinking, and emotional control. 21st century skills development, therefore, should not only concentrate on the current popular areas of STEM (Science, Technology, Engineering, Mathematics), but should also emphasize the development of communication, collaboration, creativity, and critical thinking as well.

Concluding Remarks: AI: Balancing Act

Popular media coverage on AI focuses heavily on its benefits or on its threats to humans, and seldom takes a balanced view on it. Benefits include increasing the effectiveness of diagnosis in healthcare, decreasing the cost of providing financial services to the masses, or reducing the impact of accidents due to a better flow of traffic

through autonomous driving. Risks include unjust loan decisions by banks due to biases prevalent in the data, or potential threats on humans by autonomous weapons, or governments using the immense data they collect over their citizens. Taking a single point of view of AI, focusing only on its benefits or its threats, limits the policy makers' capabilities to shape the right governance mechanism. Policy makers, therefore, need to take into consideration the interests of all parties involved, including citizens, corporations, international bodies, academia, and communities. Such a balanced view of AI would not be possible without the involvement of multiple parties working together on a common goal. A good example of a corporation taking that balanced view, for instance, would include "Ethics Committees" before the implementation of AI projects, involving key business departments, legal teams, ethics experts, technical teams. Similarly, a government acting through that balanced view would enable startups and corporations to work on AI projects, including on its citizens, but would set up an environment of transparency in data collection, processing and deletion (such as is required in GDPR). An international organization (such as UN, WHO or similar) taking that balanced view would enable the flow of data securely across boundaries, similar to how goods flow among countries. With such a balanced view, based on the pillars of Responsible AI, AI would benefit our societies immensely and would help foster democracy across the world.

THE DEMOCRATIC NEED FOR AI AS DELIBERATIVE MULTIMODAL SYSTEMS (DEMOS)

This paper explores the role and implications of AI as one of the units in the political systems, as demos, alongside people and institutions. By giving people a particular access to the word through generating textual narratives, digital personas and dialogical interfaces, AI actively participates in democratic practices. It is not, to be clear, an independent conscious agent in itself. Rather, by helping to mediate our intuitions and perceptions, as well as suggesting certain pathways for action while disclosing possible alternatives, AI merits a closer philosophical, design and policy attention in the context of democracy. The question regarding AI and democracy is, then, twofold: How does AI facilitate our practice of democracy and the way we understand it? And how can we facilitate a responsible design and use of AI systems for a meaningful democratic engagement? In this article, I will suggest that the theoretical approaches of sociotechnical systems and technological mediation can help us with both questions, while the cases of AI-mediated textual and visual input (e.g. voice assistants, GPT-3 or deepfakes) can showcase the challenges of the responsible design of AI systems for democracy.

Olya Kudina*

TPQ
Winter 2021/22

* Olya Kudina is an Assistant Professor, Ethics/Philosophy of Technology, Values, Technology and Innovation Department at Delft University of Technology.
The author would like to thank the TU Delft AI Labs initiative for facilitating the research on this paper.

Much of the recent philosophical, design and policy agendas regarding AI have been focused on *democratizing AI*, or opening the processes of AI development and decision-making. International cases of AI application in the governmental systems keep suggesting that, contrary to the hopes of more objective and efficient decision-making, AI is poorly equipped for the just allocation of tax benefits and returns,[1] fails to fairly predict the chances for recidivism in parolees [2]and promotes unfair treatment in the city[3] in an attempt to deploy better safety and surveillance regimes. It is thus not surprising that addressing the problems of algorithmic transparency, bias, fairness and accountability[4] tops the agendas of the stakeholders across the world. While very important, this article focuses on the problem that goes hand-in-hand with the first one but has to date received little attention, namely what deploying *AI for democracy* could mean. As I will argue, apart from making the AI developmental processes more transparent and fairer, it is equally important to understand how AI is subtly transforming the nature of our democratic processes, changing the values that underlie it and with that in mind, use the AI potential to strengthen our capacities for engaging with one another.

Democracy and AI

Taking cue from the political philosopher Hannah Arendt, I suggest that the value and foundation of democratic societies are in their plurality. As conditional for democracy, plurality and political activity are less concerned with seeking compromise among different parties but are primarily valued as the enactors of agency and as enablers of the human capacities for deliberation and reflective judgment.[5] Democracy here means finding a way of being together amid multiple perspectives, whereas plurality is about crafting a space for public interaction and deliberation. As we have seen above, in recent years, AI has both mediated our capacity to interact with one another and has co-shaped our infrastructure for interaction and deliberation. I would like to focus on two specific instances of how AI facilitates human capacity for critical engagement with each other, namely by affecting how we perceive and

[1] Gabriel Geiger, "How a Discriminatory Algorithm Wrongly Accused Thousands of Families of Fraud," *Vice*. 1 March 2021, https://www.vice.com/en/article/jgq35d/how-a-discriminatory-algorithm-wrongly-accused-thousands-of-families-of-fraud

[2] Karen Hao, "AI is sending people to jail—and getting it wrong," *MIT Tech Review*, 21 January 2019, https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/

[3] Dunja Mijatović, "Ethnic profiling: a persisting practice in Europe," *Council of Europe*, 9 May 2019, https://www.coe.int/en/web/artificial-intelligence/-/ethnic-profiling-a-persisting-practice-in-europe

[4] Kate Crawford and Trevor Paglen, "Excavating AI: The politics of images in machine learning training sets," *AI and Society* (2019), doi: 10.1007/s00146-021-01162-8; Andreas Holzinger, Chris Biemann, Constantinos S. Pattichis and Douglas B. Kell, "What do we need to build explainable AI systems for the medical domain?," *arXiv preprint* (2017), arXiv:1712.09923; Nicholas Diakopoulos, "Transparency." in *The Oxford Handbook of Ethics of AI*, eds. Markus D. Dubber, Frank Pasquale and Sunit Das (New York: Oxford University Press, 2020), p. 197-214.

[5] Hannah Arendt, *The Human Condition* (Chicago: University of Chicago Press, 2013. Originally published in 1958).

understand each other through language, and by helping to shape the visual aspect of reality.

> *"As conditional for democracy, plurality and political activity are less concerned with seeking compromise among different parties but are primarily valued as the enactors of agency and as enablers of the human capacities for deliberation and reflective judgment."*

Language is crucial to making sense of the everyday world, in how we understand and voice ourselves in democratic societies. Recently, an introduction of GPT-3 and similar algorithmic text-generation models, showed how AI can produce text virtually indistinguishable from that written by human.[6] While this generates excitement over potential efficiency and additional creativity, GPT-3 also raises questions for democratic participation related to the values of comprehension, verification and informed engagement when written communication is concerned.[7] The pervasive spread of voice assistants equally raises questions, now related to spoken language. While they make interaction with technologies more natural, simply by speaking to them, voice assistants also invoke a top-down command-based model of conversations and reduce the space for conversation based on language proficiency and assumed standard user capabilities[8]. The current practices with Natural Language Processing and Automatic Speech Recognition algorithms beg the question of how AI can instead promote trust, dialogue and mutual understanding, receptive to diverse positions.

Apart from language, to participate in democracy we also rely on *what we see*, driven by an 'occularcentric' bias inherent to humans. Recent developments in synthetic media complicate our tendency toward "seeing is believing," thus making us question the trust in the observed world. AI algorithms increasingly help to superimpose a face of one's choice onto specific images and videos, making the people in those videos say or do things they never have in reality (e.g. a deepfake of Queen Elizabeth[9]), recreate and manipulate 3D virtual models of a real person (e.g. Holocaust

[6] Alec Radford et al. "Better Language Models and Their Implications", *OpenAI*, (14 February 2019), https://openai.com/blog/better-language-models/#sample6

[7] Shannon Vallor, "GPT-3 and Missing Labor of Understanding." *Philosophers On GPT-3 (update with replies by GPT-3), DAILYNOUS,* 30 July 2020, https://dailynous.com/2020/07/30/philosophers-gpt-3/#vallor

[8] Olya Kudina, "'Alexa, who am I?': Voice Assistants and Hermeneutic Lemniscate as the Technologically Mediated Sense-Making," *Human Studies*, Vol. 44, No. 2 (2021): p. 233-253.

[9] "Deepfake queen to deliver Channel 4 Christmas message," *BBC News*, 23 December 2020. https://www.bbc.com/news/technology-55424730

victims "speaking" in museums[10]) or even create a fully synthetic digital being, looking as if a real person.[11] Among the abundance of ethical issues, e.g. regarding the contribution of altered media to the rise of fake news, democratic concerns take center stage. How does synthetic media affect the concepts of trust, truth and self and together, what kind of democratic practices this fosters? In our visual-biased nature, the obscured mediating qualities of synthetic media can have far-reaching consequences for how we understand each other and the world around us.

It seems that there is a new challenge for AI systems to facilitate and enhance the democratic practices, helping us to express ourselves and engage with one another. Such systems, on the one hand, need to maintain openness to the changing conditions of the world, and on the other, to maintain the core democratic values to facilitate meaningful engagement with AI and promote an informed decision-making regarding it. Answering these questions requires ethical accompaniment and the responsible design of multimodal AI systems that combine both the visual and textual modalities.

Beyond these calls for action to responsible design lurks a deeper democratic implication on the role of AI in society. I suggest that AI should be considered not only as a neutral instrument in the hands of people who design and deploy it. Rather, AI should be considered as *demos*, or one of the units in ancient societies that compiled a democratic government. Originating from Ancient Greek, *δῆμος* or *dêmos,* means common members or representatives from one area.[12] Designating AI as demos not only makes explicit the daily interactions people have with AI systems in democratic contexts but also underscores the fact that AI gives people a particular access to the engagements with others, facilitating their self-expression through the visual, written and spoken modes of activities, e.g. by generating textual narratives (e.g. GPT3), digital personas (e.g. deepfakes) and dialogical interfaces (e.g. voice assistants).

With all of these considerations in mind, the questions regarding AI and democracy seem to be twofold: *How does AI facilitate our practice of democracy and the way we understand it? And how can we facilitate a responsible design and use of AI systems for a meaningful democratic engagement?* For AI to facilitate and strengthen democratic practices, a responsible way forward seems to consist in the development of multimodal AI systems that would foster deliberation, critical engagement of citizens and enable them to take informed decisions under the conditions of deep uncertainty that these same AI developments bring forth. The multidisciplinary

[10] "Artificial intelligence project lets Holocaust survivors share their stories forever," CBS News, 3 April 2020, https://www.cbsnews.com/news/artificial-intelligence-holocaust-remembrance-60-minutes-2020-04-03/

[11] www.thispersondoesnotexist.com [The speculative and educational web project]. Accessed on 15 February 2022.

[12] The Editors of Encyclopaedia Britannica. "Deme," *Encyclopedia Britannica*, 20 July 1998. https://www.britannica.com/topic/deme-ancient-Greek-government

research and development practices,[13] spanning across at least the fields of engineering, design and ethics of technology, suggest that albeit achieving this is not an easy task, it is possible, and necessary to try.

> *"Algorithmic systems invite what gets on the political agenda and in which way people can engage with it – and as such, AI gives rise to particular ways of political agency."*

Developing AI for Democracy

From the ethics of technology side, a combination of approaches is needed that both (a) positions AI systems not as neutral instruments but as *the mediators of democratic practices* that actively participate in and co-shape such experiences[14]; and (b) that examines these AI systems not only from a technological standpoint but as complex and *dynamic sociotechnical systems*, entangling the social, technological, cultural and institutional components.[15] The relevant cases to consider how AI can foster democratic practices could span from the individual ones to more collective ones. On a micro-level, for instance, we could examine how synthetic media (e.g. deepfakes and GPT3) impacts and can foster human interrelation and trust, i.e. a precondition to deliberating with one another; or how social media algorithms co-shape an ideal of and can contribute to being a good political subject that can thrive in plurality and critically engage with others. Scaling up to the collective level, one could inquire how voice assistants and chatbots impact and can improve the representation and inclusiveness, enabling a plurality; or how algorithmic systems in the city affect political deliberation and can enlarge the space for critical and informed engagement with AI systems that support our collective life, thus facilitating active citizenship. Let us consider some of these cases in more detail.

Some of the concerns surrounding the application of synthetic media in the context of democracy center around its impact on human interrelations and trust, the trust regarding what one is seeing and hearing and the trust in others, met online. Such concerns are warranted as the online world has recently become permeated

[13] Evgeni Aizenberg and Jeroen Van Den Hoven, "Designing for human rights in AI," *Big Data & Society*, Vol. 7, No. 2 (2020): https://doi.org/10.1177/2053951720949566; Aimee van Wynsberghe, "Sustainable AI: AI for sustainability and the sustainability of AI," *AI and Ethics*, Vol. 1, No. 3 (2021): p. 213-218.

[14] Peter-Paul Verbeek, *Moralizing Technology* (Chicago: University of Chicago Press, 2011).

[15] Olya Kudina, "Bridging Privacy and Solidarity in COVID-19 Contact-tracing Apps through the Sociotechnical Systems Perspective," *Glimpse*, Vol. 22, No. 2 (2021): p. 43-54.

by AI-based technologies that manipulate the textual (e.g. GPT3), photo- and video content (e.g. deepfakes) to make it seem like people in that content say or do certain things that they might not have done originally. While such synthetic media has received negative reputation for facilitating the spread of fake news and enabling identity fraud and misrepresentation online,[16] it has also allowed for therapeutic practices, allowing to "see" the future one could have or make the past "speak" in new ways.[17] Nascent practices with synthetic media also suggest the blurring boundary between "digital" and "real," "physical" worlds, for synthetic media facilitates deep uncertainty online and offline, allowing for an unbounded ontological flexibility of everything and altering the sense of self and others. From a standpoint of philosophy and ethics, some of the crucial tensions to clarify here include how synthetic media affects interpersonal relations as people's self-image comes under pressure; and the value and role of trust as the strategy to deal with uncertainty in the everyday life. The responsible design of such AI systems that precondition democratic engagement would need to explore what is needed from us in the context of synthetic media to have good quality relations among each other. It needs to consider the development of synthetic media beyond transparency and verification methods, and additionally focus on novel media literacy strategies that deal with the limits of knowledge (e.g. the fact that even when people know they are dealing with the manipulated content, they can still be affected by it). From a responsible design standpoint, one could inquire into the public reactions towards deepfake media, reflect on possible vision-centered bias and how to facilitate interpersonal trust in the world of which deepfakes are a part. Combining both ethics of technology and responsible design perspectives would allow for a systemic analysis of AI systems and proactive attitude in the design and implementation.

Another case when considering designing AI systems that could facilitate engaging and reflective democratic practices could concern the use of social media and its relation to shaping a good political subject. Algorithmic systems invite what gets on the political agenda and in which way people can engage with it – and as such, AI gives rise to particular ways of political agency. Some argue that the incessant rich data flows and deep uncertainty regarding what one sees and reads online undermine our possibility of being together, either by suppressing dialogue[18] and ways of expressing ourselves.[19] However, such discourses are shortsighted as they consider

[16] Deborah G. Johnson and Nicholas Diakopoulos, "What to do about Deepfakes," *Communications of the ACM*, Vol. 64, No. 3 (2021): p. 33-35.

[17] Nicholas Caporusso, "Deepfakes for the Good: A beneficial application of contentious artificial intelligence technology," in *International Conference on Applied Human Factors and Ergonomics* (Cham: Springer, 2020) p. 235-241.

[18] Sherry Turkle, *Reclaiming Conversation: The Power of Talk in a Digital Age* (New York: Penguin Random House, 2015).

[19] Luciano Floridi and Massimo Chiriatti, "GPT-3: Its nature, scope, limits, and consequences," *Minds and Machines*, Vol. 30, No. 4 (2020): p. 681-694.

AI predominantly as a deterministic force only limiting democratic expression. The technological mediation lens[20] allows us to take a more nuanced look at AI systems. While AI delimits some aspects of democracy, it inevitably opens up new avenues for engagement, self-expression and promotes value change that reflects societal dynamics. Consequently, what it means to be a good citizen in the age of AI, especially with the pervasive use of social media (e.g. Facebook, TikTok, Instagram), merits closer philosophical attention. Here, one could explore how people appropriate social media to make it work for them in the democratic context and study how social media affects the responsibilities of political subjects. These findings, in turn, could contribute to the critical design of social media and foster societal debate through, e.g. design provocations by developing an add-on to the social media feeds that could prompt pausing and reflecting before posting, flag specific content and give users more control over their algorithmically-curated content.[21] The combination of ethics of technology and design fields can thus be a fruitful ground for establishing what is at stake with social media regarding being a good citizen and providing innovative ways to make people more reflective of their use of social media.

Concluding Remarks

The implementation of AI in the governing structures of society has so far witnessed an inaptitude of these technological systems to deal with the public matters in a just and fair manner. At a more individual level, AI systems give shape to the spaces of interaction that tend to promote bias, discord and aggression over plurality of positions and critical engagement with information. Not surprisingly, much of the global discussion surrounding AI concerns its democratic impact, and specifically focuses on the issues of democratizing the process of AI development: demanding that it be open, inclusive, explainable and accountable. In this paper, I have suggested to focus on the other side of the democratic impact of AI, namely exploring how the use of such systems transforms the very nature of the democratic engagement, changes its underlying values and processes and how to design better and more responsible AI systems with that in mind. In short, I have proposed in this paper to consider what AI can mean for democracy and how to coordinate the complex interdisciplinary efforts with that in mind.

Additionally, I offered to position AI as demos, a unit of participation in democratic societies, in view of the different ways in which AI already co-shapes democracy. Such a conceptualization of AI as demos allows us to foreground its mediating capacities in engaging with one another and explicitly consider the complexity of

[20] Verbeek, (2011).

[21] Tobias Rose-Stockwell, "How to Design Better Social Media: On designing social tools for society," *Medium*, 18 April 2018, https://medium.com/s/story/how-to-fix-what-social-media-has-broken-cb0b2737128

its sociotechnical impact. I have shown the way AI systems already challenge our multimodal expression capacities, for instance, by reducing the space for interaction and formulation of one's own opinions when GPT-3 language models are in play or by facilitating the default distrust in others and everything we see through an unrestricted use of deepfakes and other forms of synthetic media. Inspired by Arendt, I suggested that AI can also enhance plurality and the ability to express it, on the one hand, and promote the spaces that allow for an informed, respectful and critical exchange of opinions, on the other. Achieving this requires a comprehensive AI-for-democracy approach, powered by an interdisciplinary effort of at least ethics of technology, design and governance.

I have also suggested that such a coordinated interdisciplinary effort towards developing and using AI as deliberative multimodal systems is challenging but possible. As more formal initiatives[22] spring to address this challenge, it is also important to remember that democracy starts with an individual presenting herself to others and interacting with others. In this regard, each of us has a capacity to become aware of the role AI systems play in our daily activities, questioning their seemingly trivial role in sorting the news we see or offering us which groups to join on social media. Becoming critical and informed users of AI is something that we can do already now to contribute to the long-term quest of developing AI for democracy.

[22] DeMoS AI Lab. *Delft University of Technology*. Accessed on 15 February 2022, https://www.tudelft.nl/ai/ai-demos-lab

SETTING UP AN ETHICAL FRAMEWORK AS A FIRST STEP TO COMPREHENSIVE REGULATION OF ARTIFICIAL INTELLIGENCE TOOLS IN THE JUSTICE SYSTEM

There is an urgent need for a binding and comprehensive legal instrument which regulates the use of artificial intelligence in the field of Justice. The European Ethical Charter on the use of artificial intelligence in judicial systems and their environment provides a non-binding approach to the subject and contains five fundamental principles which must govern any artificial intelligence software used in courts or by legal professionals. The use of artificial intelligence in Law and Justice may be highly beneficial, but the perils it also entails encourage the establishment of a more robust legal framework to prevent any compromise of fundamental rights.

João Arsénio de Oliveira*

TPQ
Winter 2021/22

* João Arsénio de Oliveira is the President of the Working Group on Quality of Justice (CEPEJ) and works at the Ministry of Justice of Portugal.

The World lacks a comprehensive regulation on artificial intelligence. The judicial legal systems (international, national, regional, and local) lack a comprehensive and binding regulation on artificial intelligence applicable to Courts and the judicial environment, as a whole.

The Council of Europe, however, keeps striving to establish guidelines for the use of artificial intelligence in all fields of knowledge, as well as for the judicial systems, with the main aim (which is in line with its own goals) of avoiding any kind of infringement of the rights enshrined in the European Convention of Human Rights and, therefore, preserving the Rule of Law and Democracy. Although it is a regional international organization, its pioneer work serves as a model for other continents.

The Council of Europe decided to take a step-by-step approach. It started by establishing an Ethical Charter on the use of artificial intelligence in judicial systems and their environment[1], while, at the same time, but at a slower pace, set up a group to establish a binding instrument on artificial intelligence, which will also have provisions applicable to Courts and Law. This group was recently replaced by a new formation, whose main mission will be to draft an international instrument aimed at setting the binding rules for artificial intelligence in Europe (and, most probably, beyond).

In fact, the Council of Europe is the international organization most concerned with artificial intelligence and its implications in society, although other organizations, such as the United Nations and OECD, are also working on the subject.

One of the bodies of the Council of Europe, the European Commission for the Efficiency of Justice (CEPEJ), has been working on the issue for some time. One of its most apparent outcomes is the already alluded European Ethical Charter on Artificial Intelligence in the Legal System and its Environment. In development of the ethical framework of the Member States of the Council of Europe regarding the use of the set of sciences, theories, and techniques designed to mimic human cognition by machines, the Ethical Charter forms the cornerstone of its design.[2]

The Charter is intended for public and private stakeholders responsible for the design and deployment of artificial intelligence tools and services that involve the processing of judicial decisions and data (machine learning or any other methods deriving from data science).[3] It also concerns public decision-makers in charge of

[1] Which can be found on https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment (February 2022).

[2] João Arsénio de OLIVEIRA, « La Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires, » *in Revue Experts* n° 141, (Décembre 2018), p. 30.

[3] CEPEJ, *The European Ethical Charter of the CEPEJ*, (2019): p. 5.
https://experts-institute.eu/en/europe-of-justice/cepej-en/the-european-ethical-charter-of-the-cepej/

the legislative or regulatory framework, of the development, audit or use of such tools and services.[4]

The Charter consists of five major principles and four appendices. A detailed study of the use of artificial intelligence in judicial systems is presented in the first appendix. In the second one, using artificial intelligence for its intended uses is encouraged, but only for those requiring considerable methodological precautions. This is a set of uses that should be considered after further research. An appendix contains a list of uses that should be regarded with the greatest caution; a glossary forms the third appendix and the final annex contains a checklist for integrating the Charter's principles into processing methods.

> *"The principle of quality and security requires that automatic learning can be performed on certified originals and that the integrity of these data is ensured at all stages of processing."*

The five principles enshrined in the Charter are:

1. The Principle of Respect for Fundamental Rights: ensure that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights;

2. The Principle of Non-Discrimination: specifically prevent the development or intensification of any forms of discrimination between individuals or groups of individuals;

3. The Principle of Quality and Security: process judicial decisions and data using certified sources and intangible data derived from models designed in a multidisciplinary manner, within a technologically secure environment;

4. The Principle of Transparency, Impartiality and Fairness: make data processing methods accessible and understandable, authorize external audits;

5. The Principle «Under User Control»: preclude a prescriptive approach and ensure that users are informed actors and in control of their choices.

[4] CEPEJ, (2019).

The first principle stipulates that, human rights must be respected from the inception of any artificial intelligence instrument (human rights by design). It intends to make sure that all artificial intelligence solutions respect the rights foreseen in the European Convention of Human Rights, as well as other relevant Council of Europe instruments.

Within the *acquis* of fundamental rights foreseen in the European Convention of Human Rights, the CEPEJ considered necessary to highpoint the principle of non-discrimination, since it is extremely susceptible to be breached by artificial intelligence tools in the field of Justice (mainly in criminal matters), as is better explained below.

The principle of quality and security requires that automatic learning can be performed on certified originals and that the integrity of these data is ensured at all stages of processing. The principle also encourages a multidisciplinary approach, urging the designers of machine learning models to rely on the expertise of the relevant justice system professionals and researchers in the fields of Law and social sciences. Moreover, it requires that data must come from certified sources and that it should not be modified until it has been used.

The principle of transparency, impartiality and fairness indicates that it is necessary to find a balance between the intellectual property of the processing methods and the need for transparency (access to the design process), impartiality (absence of bias), fairness and intellectual integrity (prioritizing the interests of Justice).

Last but not least, the principle 'under user control' emphasizes the importance of the human being and affirms that all artificial intelligence tools should focus on the person above all in the field of justice. Corollaries of this principle determine that, legal professionals should, at any moment, be able to review judicial decisions taken by IT and that the subjects of artificial intelligence decisions should be conferred special information rights. At the same time, the knowledge and understanding of AI should be promoted in government institutions, independent oversight bodies, national human rights structures, the judiciary and law enforcement, with addition to the general public.[5]

CEPEJ's Plenary Meeting approved a revised roadmap in December 2021 for ensuring an appropriate follow-up to the CEPEJ Ethical Charter, which established (again) five key elements aimed at increased application of the Charter.

[5] Commissioner for Human Rights, *Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights* (Recommendation) (Strasbourg: Council of Europe, May 2019): p. 14.

The first one comprises developing more practical guidance to legal professionals and IT developers on how to apply the five principles laid down in the CEPEJ Charter, through the detailed operationalization of the five principles. This means developing a detailed description of what needs to be checked and how to ensure compliance with the respective principle.

The second idea consists of creating a pilot project on CEPEJ Charter conformity assessment. It engages the operationalization of the checklist attached to the Charter on one artificial intelligence program, through a test case led by the CEPEJ secretariat.

"Even though artificial intelligence is promising (but also a threat) to all fields of law, it seems to pose a greater threat to criminal and criminal procedure law in particular."

Furthermore, the CEPEJ has decided to create a five member Artificial Intelligence Advisory Board (AIAB) to monitor the emergence of artificial intelligence applications in the justice system, to discuss issues, and to propose new ways to protect fundamental rights when using artificial intelligence. The Advisory Board will meet virtually every trimester.

The Advisory Board will also be in charge of feeding the Resource Centre on Artificial Intelligence, which will be a publicly accessible database containing reliable and exhaustive overview of the existing artificial intelligence programs in the judicial area.

The final vector of the roadmap relies on the need to increase the awareness of the CEPEJ Charter, by means of concrete training activities.

Even though artificial intelligence is promising (but also a threat) to all fields of law, it seems to pose a greater threat to criminal and criminal procedure law in particular. Several experiences have already highlighted how dangerous automatic interventions can be in the penal realm, and how aggressive they can be towards human rights. By itself, this justifies establishing a binding legal instrument applicable to the use of artificial intelligence in the judicial process.

Not long ago, a report from ProPublica[6] described that an artificial intelligence tool used in courtrooms across the United States to predict future crimes, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), was biased against black defendants.[7]

Often, the problem relies on the data used to feed the artificial intelligence machine, as it can only proceed if supplied with data. If the data fed to the machine already comprises biases, the machine and the algorithm on which it relies will only perpetuate the discriminatory behaviour and, therefore, contribute to aggravate the discrimination.

According to a recent article, researchers in China have developed a machine, using artificial intelligence that can charge people with crimes. This artificial intelligence «prosecutor» can file a charge with more than 97 percent accuracy, based on a verbal description of the case, according to the same source.[8] The lack of human intervention is scary and, as one Prosecutor interviewed by the same reporter has stated an AI prosecutor could file a charge based only on its previous experience. It could not foresee the public reaction to a case in a changing social environment. AI may help detect a mistake, but it cannot replace humans in making a decision.

Fundamental rights such as freedom of expression and assembly, right to privacy, as well as non-discrimination are at high risk of being violated by artificial intelligence tools. For this reason, we thought it fundamental to have a binding (hopefully global) instrument that regulates the use of artificial intelligence in the justice system.

[6] ProPublica is, according to its own words "An Independent, Non-Profit Newsroom that Produces Investigative Journalism with Moral Force," https://www.propublica.org

[7] Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias," *ProPublica*, 23 May 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[8] Stephen Chen, "Chinese Scientists Develop AI 'Prosecutor' that can press its own Charges," *South China Morning Post*, 26 December 2021, https://www.scmp.com/news/china/science/article/3160997/chinese-scientists-develop-ai-prosecutor-can-press-its-own

UNPACKING THE EU PROPOSAL FOR AN AI ACT: IMPLICATIONS FOR AI SYSTEMS USED IN THE CONTEXT OF MIGRATION, ASYLUM AND BORDER CONTROL MANAGEMENT

On 21 April 2021, the European Commission presented its long-awaited proposal for a Regulation laying down harmonized rules on AI.[**] The proposal, which is based on Articles 16 and 114 of the TFEU on data protection and the internal market respectively, aims at a risk-based approach that respects EU values and protects personal data and fundamental rights at the same time. This contribution aims to critically examine which AI tools fall within the scope of AI Act, how these are classified and whether the proposed approach poses concerns for the protection of fundamental rights. To that end, the next section provides a snapshot of the proposal to inform the subsequent analysis, followed by an analysis on which migration-related AI systems aiming to identify potential loopholes. In addition to an appraisal of which AI systems are envisaged within the scope of the proposal, this article assesses Article 83 of the proposal, which excludes several AI initiatives from its scope.

Niovi Vavoula*

TPQ
Winter 2021/22

* Dr. Niovi Vavoula is a lecturer in Migration and Security at University of London.
**Commission, Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (Proposal for AI Act) COM (2021) 206 final.

Overview of the Proposal for an AI Act

The proposed Regulation distinguishes AI systems on the basis of the risk they pose to the fundamental rights of individuals or EU values. In light of this, according to Article 5 of the proposal, particularly harmful AI practices should be prohibited, such as those that can manipulate vulnerable individuals through subliminal techniques. These practices may cause harm to the manipulated individual or others. For the purpose of law enforcement, social scoring and real-time remote biometric identification are not permitted in publicly accessible areas due to the 'unacceptable risk' they create.[1] 'High risk' AI systems impact people's safety or fundamental rights, and are therefore considered high-risk. Systems listed in Annex III include those used for biometric identification and categorisation of individuals, critical infrastructures (such as transportation), formal education or vocational training, essential private and public services (such as credit scoring which denies citizens a loan), law enforcement, migration, asylum, border control, management and administration of justice and democratic processes. The deployment of 'high-risk AI' systems will need to undergo a conformity assessment before being placed on the market and comply with a range of safety requirements (regarding, for instance, risk management, human oversight and data governance).[2] In addition, an *ex-post* market surveillance and supervision must be put in place to ensure compliance with the obligations and requirements for all high-risk AI systems already placed on the market.[3]

Zooming into Migration, Asylum and Border Control Management

Under migration, asylum and border management, Annex III details four different types of AI systems. These are:

(a) AI systems intended to be used by competent public authorities as polygraphs and similar tools or to detect the emotional state of a natural person;

(b) AI systems intended to be used by competent public authorities to assess a risk, including a security risk, a risk of irregular immigration, or a health risk, posed by a natural person who intends to enter or has entered into the territory of a Member State;

(c) AI systems intended to be used by competent public authorities for the verification of the authenticity of travel documents and supporting documentation of natural persons and detect non-authentic documents by checking their security features; and

[1] For criticism about the see Center for Artificial Intelligence and Digital Policy, *Center for AI & Digital Policy (CAIDP) Statement on Proposed EU AI Regulation* (28 July 2021).

[2] Center for Artificial Intelligence and Digital Policy, (2021), Chapter 2 and Article 43.

[3] Center for Artificial Intelligence and Digital Policy, (2021), Article 61.

(d) AI systems intended to assist competent public authorities for the examination of applications for asylum, visa and residence permits and associated complaints with regard to the eligibility of the natural persons applying for a status.

"Similarly, data provided by short or long-stay visa and residence permits applicants will be subject to comparable rules, except that the automated processing will never lead to automatic issuance of a visa or a residence permit and there is always manual processing of the application by visa authorities."

The first type essentially refers to emotion recognition systems that claim to infer people's emotions and mental states from physical, physiological, behavioural, as well as biometric data. Their use worldwide is currently in its early stages, but not prohibited. In the U.S, the Department of Homeland Security funded research of the virtual border agent technology known as the Automated Virtual Agent for Truth Assessments in Real-Time (AVATAR) and allowed it to be tested it at the U.S-Mexico border on travelers who volunteered to participate.[4] Perhaps the most prominent application of emotion recognition in the EU was Frontex's highly contentious iBorderCtrl project (Intelligent Portable Border Control System) from September 2016 through August 2019. Detecting mental states and emotions by examining a person's facial expressions as well as other physiognomic indicators has attracted considerable attention. Hungary, Latvia, and Greece piloted an automated lie-detection test analysing facial micro-gestures at three undisclosed airports in order to identify 'biomarkers of deceit'.[5] By scanning refugees' and migrants' facial expressions as they answer questions, the machine is able to determine whether they have lied to the machine, and then pass on that information to border officers. The project has been challenged before the CJEU on transparency grounds. MEP Patrick Breyer has requested that the EU Research Agency (REA) release classified documents evaluating the 4.5 million euro trial of AI lie detectors to bolster EU border control to determine its ethical and legal justifiability, as well as the results of the technology, on the ethical justification and legal admissibility.[6] In December 2021, the General Court of Justice opined that REA may no longer keep these documents

[4] Jeff Daniel, "Lie-detecting Computer Kiosks Equipped with Artificial Intelligence look like the Future of Border Security," *CNBC*, 15 May 2018, https://www.cnbc.com/2018/05/15/lie-detectors-with-artificial-intelligence-are-future-of-border-security.html accessed on 25 February 2022.

[5] Javier Sánchez-Monedero and Lina Dencik, "The Politics of Deceptive Borders: 'Biomarkers of Deceit', and the Case of iBorderCtrl," *Information, Communication & Society*, Vol 1 (2020).

[6] Zach Campbell, Caitlin L Chandler and Chris Jones, "Sci-fi Surveillance: Europe's Secretive Push into Biometric Technology," *The Guardian*, 10 December 2020, https://www.theguardian.com/world/2020/dec/10/sci-fi-surveillance-europes-secretive-push-into-biometric-technology accessed on 25 February 2022.

completely secret. The ethical and legal evaluation of technologies for 'automated deception detection' or automated risk assessment must be published, as long as they do not relate specifically to the iBorderCtrl project.[7] Nevertheless, to protect commercial interests, the examination of ethical risks (e.g. stigmatization and false positives) and legal admissibility of the concrete technology of iBorderCtrl, as well as reports of project results, may be kept confidential.

Considering that the EU has already been testing AI polygraphs and lie detectors for years, it is unsurprising that the proposal does not wish to ban them entirely. This is notwithstanding the fact that the programme has reported an accuracy rate of 73-75 percent and has raised (well-deserved) criticism about its validity that its applicability in a real-life context.[8] Similarly, the U.S. AVATAR as a deception-detection judge has a success rate of 60 to 75 percent and sometimes up to 80 percent.[9] Expressions and facial expressions differ greatly between cultures and situations. As a result, emotion detection has limited reliability;[10] emotion categories are neither reliably expressed through, nor unequivocally associated with a common set of facial movements. Furthermore, facial expressions do not perfectly match emotion categories, so generalizability is generally limited. In addition, the effectiveness of such tools has yet to be proved, whereas the risk of racial profiling is significant.[11] From that perspective, the judgment by the General Court is somewhat disappointing, as the results of the project could support the claim that emotion detection systems posed such ethical and fundamental rights challenges to the extent that they should not be classified as high risk and should be banned altogether, as advocated by civil society actors.[12] At least, it is obligatory for project participants to make a scientific publication about the project within four years and therefore it is possible to get more answers on the project.

Whenever automated assessments are made as to whether an individual may pose a security, immigration, or health risk, these are rightly classified as high-risk AI

[7] Case T-158/19 *Patrick Breyer v European Research Executive Agency*, (2021), ECLI:EU:T:2021:902.

[8] Javier Sánchez-Monedero and Lina Dencik, (2020).

[9] Daniel, 15 May 2018.

[10] It has even been called as 'pseudoscience' because the 'micro-expressions' the software analyses cannot be reliably used to judge whether someone is lying. In 2019 The Intercept tested the iBorderCtrl system. The video lie detector falsely accused its reporter of lying — judging she had given four false answers out of 16, and giving her an overall score of 48, which it reported that a policeman who assessed the results said triggered a suggestion from the system she should be subject to further checks (though was not as the system was never run for real during border tests). See Ryan Gallagher and Ludovica Jona, "We Tested Europe's New Lie Detector for Travelers – And Immediately Triggered a False Positive," *The Intercept*, 26 July 2019, https://theintercept.com/2019/07/26/europe-border-control-ai-lie-detector/ accessed on 25 February 2022.

[11] American Civil Liberties Union (ACLU), *Bad Trip: Debunking the TSA's 'Behavior Detection' Program* (2017).

[12] "An EU Artificial Intelligence Act for Fundamental Rights a Civil Society Statement," *Accessnow*, 30 November 2021, https://www.accessnow.org/cms/assets/uploads/2021/11/joint-statement-EU-AIA.pdf accessed on 25 February 2022.

"Chatbots use natural language processing to conduct human-like conversations with users and they have been criticised for embedding biases and using discriminatory language. Therefore, these should also be considered as high risk."

systems. A subset of automatic assessments falls under this category of AI, which incorporate algorithmic profiling against risk indicators developed through statistical analysis of rates of overstaying and refusals of entry. It also includes the use of AI tools to identify irregular traveling patterns as an additional piece of risk analysis and identify so-called 'mala fide' travelers. This category also includes AI systems that determine whether an asylum seeker poses a risk of absconding and should therefore be detained. The EU legislature has already adopted legal instruments that will require in the near future such automated assessments in the cases of the European Travel Information and Authorisation System (ETIAS)[13] and the Visa Information System (VIS).[14] Visa-free nationals will have to register for ETIAS online in order to provide a wide array of personal data, which will then be cross-checked against databases, both European and some Interpol ones, according to specific screening rules, in order to enable profiling on the basis of risk indicators. An ETIAS authorization will be automatically provided in the event that there is no problem with the data provided, otherwise the application will go through manual processing by the responsible Member State. Similarly, data provided by short or long-stay visa and residence permits applicants will be subject to comparable rules, except that the automated processing will never lead to automatic issuance of a visa or a residence permit and there is always manual processing of the application by visa authorities.[15] As argued elsewhere, the risk of discrimination both direct and indirect in the design of the algorithms is particularly high. As they are trained using pre-existing data and past decisions, they run the risk of repeating all of the implicit and inherent biases of those earlier decisions. Such initiatives are already tried (unsuccessfully)

[13] Regulation (EU) 2018/1240 of the European Parliament and of the Council of 12 September 2018: *Establishing a European Travel Information and Authorisation System (ETIAS) and Amending Regulations (EU) No 1077/2011*, (*EU) No 515/2014, (EU) 2016/399, (EU) 2016/1624 and (EU) 2017/2226* [2018] OJ L236/1.

[14] Regulation (EU) 2021/1134 of the European Parliament and of the Council of 7 July 2021 *Amending Regulations (EC) No 767/2008, (EC) No 810/2009, (EU) 2016/399, (EU) 2017/2226, (EU) 2018/1240, (EU) 2018/1860, (EU) 2018/1861, (EU) 2019/817 and (EU) 2019/1896* of the European Parliament and of the Council and repealing *Council Decisions 2004/512/EC and 2008/633/JHA*, for the purpose of reforming the Visa Information System [2021] OJ L248/11.

[15] For further information see Niovi Vavoula, "Artificial Intelligence (AI) at Schengen Borders: Automated Processing, Algorithmic Profiling and Facial Recognition in the Era of Techno-Solutionism," *European Journal of Migration and Law*, Vol. 23, No. 4 (2021): p. 457.

in the UK, where the Home Secretary decided to withdraw an algorithm used to filter visa applications after being called out as racially discriminatory.[16] The system took some information provided by visa applicants and automatically processed it, giving each person a colour code based on a "traffic light" system - green, amber, or red. Similarly, in Canada the Government tested AI solutions to triage temporary resident visa applications.[17] Regrettably, these systems will not fall within the scope of the AI Act. According to article 83, AI systems which are components of large-scale IT systems governed by a series of EU legal instruments (listed in Annex IX) and implemented before 12 months after the application date of the AI Act - after its enactment - are excluded from the AI Act, unless the replacement or amendment of those legal acts results in a significant change to the design or intended purpose of the AI system or AI systems concerned. As ETIAS is slated for implementation in 2023, and the VIS rules will be in place not long after, it is almost certain that algorithmic profiling of travelers and migrants will fall outside the scope of the AI Act, which is still being negotiated and will take two years to enforce in the Member States. That said, the principle that automated risk assessments should be classified as high risk will still be applicable, but the deployment of AI systems in relation to these databases will not be subject to the safeguards encompassed in the forthcoming AI Act.[18]

The exclusion clause of Article 83 is particularly problematic for a series of other reasons. In addition to ETIAS and VIS, the remaining information systems that process personal data of third-country nationals, Schengen Information System (SIS), Eurodac, Entry/Exit System and European Criminal Record Information System for Third-Country Nationals (ECRIS-TCN) are also under development

[16] Henry McDonald, "Home Office to scrap 'racist algorithm' for UK visa applicants," *The Guardian*, 4 August 2020, https://www.theguardian.com/uk-news/2020/aug/04/home-office-to-scrap-racistalgorithm-for-uk-visa-applicants

[17] Petra Molnar and Lex Gill, "Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System," (2018); Lucia Nalbandian, "Using Machine-Learning to Triage Canada's Temporary Resident Visa Applications," (Working Paper 2021/09, July 2021).

[18] The requirements however prescribed in Opinion 1/15 and *La Quadrature du Net and Others* about automated analysis will still be applicable. In particular, according to Opinion 1/15, a) pre-established models and criteria (algorithms), should be 'specific and reliable' to individuals who might be under a 'reasonable suspicion' of participation in terrorist offences or serious transnational crime and should be non-discriminatory; b) that the databases cross-checked must be reliable and up to date; c) any hit following the automated processing of that data must be subject to an individual re-examination by non-automated means; d) the pre-established models and criteria and the databases used are not discriminatory and are limited to that which is strictly necessary, the reliability and topicality of those pre-established models and criteria and databases used should, taking account of statistical data and results of international research. In *La Quadrature du Net and Others*, the Court of Justice of the EU added that the models or criteria to conduct an automated analysis cannot be based on sensitive data *in isolation.* Furthermore, the Court acknowledged the need to regularly re-examine the algorithms and the databases used to ensure that they are reliable, up-to-date and in practice non-discriminatory and limited to what is strictly necessary to achieve the intended purpose. Lastly, the Court held that because of the margin of error that may result from the automated analysis, there has to be an individual non-automated re-examination of a positive result before adopting a measure that may have adverse effect on the person concerned. See Opinion 1/15 ECLI:EU:C:2017:592; Joined Cases C511/18, C512/18 and C520/18 *La Quadrature du Net and Others v Premier Ministre and Others* ECLI:EU:C:2020:791.

(EES and ECRIS-TCN) or refurbishment (SIS and Eurodac) to encompass among others facial recognition technology, which enables biometric identification and is also considered as high risk type of AI systems. Consequently, Article 83 essentially excludes all information systems for third-country nationals from the protective scope of the AI Act. This exclusion applies unless those systems are subject to 'significant changes' in design or intended purpose. As mentioned in the Joint Opinion of the European Data Protection Supervisor (EDPS) and the European Data Protection Board (EDPB), the threshold for 'significant changes' is not clear. Recital 66 of the Proposal specifies a lower threshold for conformity re-assessment 'whenever a change occurs which may affect the compliance' and it has been rightly argued, a similar threshold would be appropriate for Article 83 as well, at least for high risk AI systems.[19] According to the Commission report on the opportunities and challenges for the use of AI in border control, migration and security, additional initiatives are underway, including application triaging, along the lines of the UK failed experiment. [20] In the case of ETIAS, the report notes that there will be in the future individual risk assessment by an AI tool to assist in manual processing.[21] These initiatives should be considered as 'significant changes' to the functioning of the systems to subject them to the forthcoming AI Act. Therefore, if these initiatives are to be adopted and implemented within the next few years, it makes no sense to keep the exclusion clause. Overall, it is unclear why this exclusion clause, which will create protection gaps, has been included in the first place. One possible explanation is the fact that the AI systems that will be included in the information systems for third-country nationals are already under development and awaiting for the adoption of the AI Act could potentially change the timeline for implementation as well as change the procedures for testing the algorithms that will be used, which are to be developed by the ETIAS Central Unit, within the European Border and Coast Guard. In any case, and as a *minimum*, considering that the entry into application is envisaged for 24 months following the entry into force of the future Regulation, exempting AI systems already placed on the market for an even longer period of time is not appropriate.[22] Moreover, while the Proposal also provides that the requirements of the Regulation shall be taken into account when evaluating each large-scale IT system as provided by the legal acts listed in Annex IX, both EDPB and EDPS believe that conditions for putting AI systems into service should be applicable from the date of application of the future Regulation. That said, Article 83 contains a protection clause; the requirements of the AI Act must be taken into

[19] EDPS and EDPB, *EDPB-EDPS Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence* (Artificial Intelligence Act) (18 June 2021): p. 14.

[20] Commission, "Opportunities and Challenges for the Use of Artificial Intelligence in Border Control, Migration and Security," (May 2020): p. 16.

[21] Commission, (May 2020): p. 19.

[22] EDPS and EDPB, (18 June 2021): p. 14.

account, where applicable, in the evaluation of large-scale IT systems. This is an important safeguard, but it must be noted that in the past, the legal bases of information systems have been amended before the conduct of an evaluation and therefore it is possible that the AI Act prescriptions will not be taken into account at that phase. Thus, it is imperative that the threshold of placing existing AI systems within the scope of the Act is lowered to minimize gaps in protection.

Under the third type of AI systems, AI systems might enable tools that detect forgeries of travel documents or other supporting documents. Furthermore, these systems are rightly regarded as high risk, since they are bound to significantly affect individuals if the results (false positives and negatives) are inaccurate. For example, various unjustified claims could be made against an individual (e.g., document fraud) on the basis of inaccurate underlying data. Meanwhile, individuals may face detention and degrading treatment when suspected of criminal conduct.

In regard to the last type of AI systems classified as high risk, the proposal refers to assistance in examining requests for asylum, visas, and residence permits, as well as complaints regarding eligibility of the natural person applying for such a status. AI systems of this type seem closely related to the second, but they may encompass broader initiatives such as personalized application forms that tailor questions to the applicant using AI. In this case, augmented application forms could be developed, or artificial intelligence systems could be used to perform vulnerability assessments of asylum seekers to determine if they need to be further investigated by a social worker or granted special procedural rights. The repercussions of severe decisions and actions in this context can have particularly severe consequences on individuals; the remedies to reverse the harm caused by these actions and decisions are unclear.

Finally, there are three further reasons why the list of migration-related AI systems is problematic. The first one concerns the lack of references to the use of AI systems for predictive analytics of migration.[23] As mentioned by Beduschi, AI technologies to foresee arrivals and prepare more efficiently for large influxes of people.[24] In February 2022, the Commission along with the European Union Asylum Agency (former European Asylum Support Office) announced the development of a forecasting model for asylum flows.[25] In addition, the use of chatbots and intelligent

[23] European Digital Rights (EDRi) and others, "An EU Artificial Intelligence Act for Fundamental Rights A Civil Society Statement," *Accessnow*, (30 November 2021). https://www.accessnow.org/cms/assets/uploads/2021/11/joint-statement-EU-AIA.pdf accessed 25 February 2022

[24] Ana Beduschi, "International Migration Management in the Age of Artificial Intelligence," *Migration Studies*, Vol. 9, No. 3 (2020): p. 576.

[25] Marcello Carammia, Stefano Maria Iacus and TeddyWilkin, "Forecasting Asylum-related Migration Flows with Machine Learning and Data at Scale," *Nature Scientific Reports* (2022). Prior to this, EASO developed an Early Warning and Forecasting System that allows for the prediction of migration movements, which used as a basis social media data, which was outside of its mandate and for which the agency was criticised.

agents, which would take in information, answer questions posed by the applicant, and ensure data quality, also falls outside the scope of the AI Act. Chatbots use natural language processing to conduct human-like conversations with users and they have been criticized for embedding biases and using discriminatory language.[26] Therefore, these should also be considered as high risk. Furthermore, individuals should know in advance whether they are dealing with a human or a machine. Finally, the proposal refers to AI systems used by public authorities only, whereas increasingly there are instances of border management entrusted to private companies, for example, in the management of camps in the Greek islands.[27] Due to the blurred lines between public and private security it was imperative that the use of AI system for immigration control purposes would fall within the scope of the AI Act regardless of the actor involved. In fact, the latest Council documents consolidating the amendments proposed during the Slovenian presidency address this concern.[28]

This contribution pointed out that the proposed EU AI Act categorizes as high risk several AI systems that may be used for immigration-related purposes; however, there are significant flaws in the current approach, which the Council has only partially rectified. Negotiations are still ongoing, so it will be up to the Parliament to push for a better balance between taking advantage of the opportunities offered by AI and protecting fundamental rights.

[26] Christopher Heine, "Microsoft's Chatbot 'Tay' Just Went on a Racist, Misogynistic, Anti-Semitic Tirade," *AdWeek*, 24 March 2016, https://www.adweek.com/performance-marketing/microsofts-chatbot-tay-just-went-racist-misogynistic-anti-semitic-tirade-170400/ accessed on 25 February 2022.

[27] "Concerns over States contracting Private Security Companies in Migration Situations," (United Nations High Commissioner for Human Rights, (20 December 2019). https://www.ohchr.org/EN/NewsEvents/Pages/SecurityPrivatisationMigrationContexts.aspx accessed on 25 February 2022.

[28] Council, *Document 14278/21* (29 November 2021).

HOW REFRAMING PRIVACY WOULD UPHOLD DEMOCRACY IN THE DIGITAL AGE

AI and algorithmic decision making are everywhere, and increasingly determine what happens to us, as well as editing the news and adverts we see, acting as our gateway to reality. The world is grappling with these consequences of AI. Many countries, businesses and organizations are equipping themselves with policies, guidelines and strategies to harness the benefits of this technology while tempering the risks. I argue that privacy can continue to make an effective and substantial contribution to upholding human rights and democracy in our datafied societies – but this can only happen if we reframe privacy as a public concept, making privacy law more dynamic on the way.

Ivana Bartoletti*

TPQ
Winter 2021/22

* Ivana Bartoletti is the Global Chief Privacy Officer at Wipro Technologies, and the Founder of the Women Leading in AI Network. Author of An Artificial Revolution: on Power, Politics and AI (Indigo Press), Ivana is a Visiting Policy Fellow at the University of Oxford.

The birth of the internet promised to bring freedom, to connect people globally, break barriers and help achieve liberty, democracy and equality. However, the digital ecosystem that we now inhabit does not deliver on those original ideals.

We live with the consequences of political inaction around digital governance, including weaponized social media, cyber-intrusions that prey on the vulnerabilities of internet architecture, the distorting marketplace of AI-informed predictions about individual internet users' future behaviours, and information monopolies that threaten democratic discourse online.

The good news of the last year is that these negative results have started to be challenged by countries and blocs around the globe. On 21 April 2021, the European Commission presented a proposal for a Regulation concerning artificial intelligence (AI), – the AI Act, for short. This draft AI Act seeks to lay down harmonised rules for the development, placement on the market and use of AI systems. It takes account of their widely varying characteristics and risks, including outright prohibitions and a more nuanced conformity assessment system adapted from EU product safety law. This recognizes that while AI offers us all tremendous opportunities, it carries risks for the fundamental rights of individuals.

In parallel, the EU is proposing a full range of legislative tools aiming to address market dominance and rein in the power of large tech companies. A similar approach is being taken in other countries, from the U.S. to China.[1] For example, Beijing has clipped the wings of its once seemingly untouchable technology giants in a dramatic clash between public and private power and, despite the potentially chilling effects on innovation and economic growth, China's regulators look set to continue high profile enforcement actions in 2022, furthering President Xi Jinping's bid for "common prosperity."

Meanwhile in the U.S., the FTC (Federal Trade Commission) has underscored the links between the protection of privacy and competition law and, under the leadership of its chair, Lina Khan, is enforcing antitrust law and investigating companies.

At the same time, we have witnessed a few years of the exposure of data mishandling and the unfair treatment of women and P.O.C. (People of Colour) in tech, leading to unprecedented levels of tech activism. Unionization of tech workers has surged alongside the awareness that technology is far from neutral: it is embedded into the existing structures of our societies and is very efficient at entrenching existing

[1] Brian Liu and Raquel Leslie, "China's Tech Crackdown: A Year-in-Review," *Lawfare*, 7 January 2022, available at: https://www.lawfareblog.com/chinas-tech-crackdown-year-review (last accessed: 20 February 2022).

inequalities and indeed escalating them.

Within this context, 'ethics' became the new buzzword: over 180 different sets of guidelines have been published to try and define and implement the key principles when developing a fair, responsible and trustworthy use of AI.

My own view is that the debate around ethics has been positive. The incredible contributions of superb advocates including Joy Boulanwini have been crucial to bring these issues into the mainstream. Facebook's decision to delete one billion "faceprints" of real people (used as part of a facial recognition system for photo tagging) would not have happened without the relentless and powerful work of Boulanwini and numerous others.

> *"As AI systems increasingly place allocative functions in our society (on whether we receive a loan, for example), it appears increasingly complex to conceive privacy merely as a set of individual rights that one is entitled to exercise."*

Many in this impressive army of committed ethicists will share my concern about how ethics programmes are being swallowed up by existing tech company mind-sets and practices, and too often just subsumed with a smile. The key point is that ´when ethical ideals are at odds with a company's bottom line, they are met with resistance'[2]. Furthermore, even if individuals genuinely want to create impactful and meaningful change, companies seem adept at taking ethics under their corporate affairs mantle, without properly examining let alone addressing their business strategies and processes.

This is why I argue that privacy and data protection law can make a significant contribution to how we uphold human and social rights in the age of AI and algorithmic decision making. Now more than ever, I am convinced we need to reflect on how privacy, as currently conceived, is tested by big data and AI. In this vein, I hold that the pervasiveness of AI, profiling and algorithmic functions require us to adapt existing legislation in a more progressive and aspirational way.

[2] Andrew Marantz, "Silicon Valley's Crisis of Conscience", *The New Yorker*, 26 August 2019. https://www.newyorker.com/magazine/2019/08/26/silicon-valleys-crisis-of-conscience (last accessed: 20 February 2022).

The Importance of Data Accuracy

In July last year, *Foodinho*, a subsidiary of GlovoApp23, was fined EUR 2.6 million by the Italian Supervisory Authority (SA) known as *Garante per la Protezione dei Dati Personali.* Foodinho was required to amend the way it processes its riders ('delivery drivers') data through a digital platform. On top of that, it was required to verify that the algorithms used to book and assign orders for food and other products do not result in discrimination.

In a joint operation by their supervisory authorities, Italy and Spain (through its AEPD) investigated the digital platform owned by *GlovoApp23*, as the holding company. They identified several infringements of privacy law, particularly in how the algorithms were used to handle employees. For instance, the company had failed to adequately inform its employees on the functioning of the system they were subject to; and it had not implemented suitable safeguards to ensure that the algorithmic results that were used to rate riders' performance were accurate and fair. Nor did the company have any procedures in place to enforce the right to obtain human intervention, to express one's point of view and contest the automated decisions taken using those algorithms – which in some cases excluded a rider from taking up assignments they wanted to work on.

Foodinho was ordered to check the accuracy and relevance of the data used by the system – chats, emails and phone calls between riders and customer care, geolocation at 15-second intervals, mapping of routes, estimated and actual delivery time, details on the handling of current and past orders, feedback from customers and partners, device battery level, etc. This order was also intended to minimize the risk of errors and biases that might for instance result in reducing delivery assignments to certain riders or excluding a rider altogether from the platform. Such risks are also related to the company's rating system, which, relying on the application of a mathematical formula, dished out penalties for riders who did not accept orders promptly enough, or rejected the orders, whilst rewarding riders who accepted orders swiftly or delivered the most orders. The rating takes into account delivered orders, check-ins performed within each booked slot a few minutes after the start of the slot and acceptance of the assigned order *within 30 seconds*.

The above is a clear example of a regulator leveraging existing data protection legislation in an algorithmic context.

The Limits of the Rights Argument in the Context of AI

The ubiquity of data collection poses challenges from a privacy and data protection

standpoint, and we now live in a fully datafied society. Datafication of our persons takes place through various systems that authenticate our identity, allowing us to fulfill our duties (e.g. to file an income tax return) or benefit from our rights (e.g. to enjoy family life, or to receive child benefits or housing subsidies)

So our personhood also expresses itself through a datafied identity.

The key question is how the existing conceptualization of privacy can cater for our datafied identities.

"The current draft EU AI Act proposes self assessments for high risk AI, except for facial recognition technology (FRT), following the product safety legislation already adopted in the EU through the CE marks."

As Daniel Solove[3] puts it, there are two dimensions to data. The first is that data is shared between people. Our datafed identities interact with each other. As Simeon de Brouwer observes, there are "privacy externalities" because people's decisions to allow the collection and use of their personal data can also reveal data about other people[4].

As Solove continues, ´rights also are limited when people's privacy decisions involve interrelated data. Interrelated data is somewhat different than shared data. Shared data involves facts that are directly connected between two or more people – such as genetic data. Interrelated data involves data that is not necessarily shared but that affects inferences made about others. In today's "inference economy," machine learning and other forms of algorithmic decision making work by making inferences based on data sets. Everyone's data in the data set is used to make inferences, which are often then used to make decisions affecting people. As Salomé Viljoen notes, "Data flows are designed to represent the ways that people are like one another and reveal meaningful things about one another; how we are alike biologically, interpersonally, politically, and economically. In sum, because a person's data might be the piece in a puzzle that also enables inferences to be made about others, that

[3] Daniel Solove, *The Limitations of Privacy Rights*, 1 February 2022, available at *SSRN*: https://ssrn.com/abstract=4024790 or http://dx.doi.org/10.2139/ssrn.4024790 (last accessed: 20 February 2022).

[4] Simeon de Brouwer, "Privacy Self-Management and the Issue of Privacy Externalities: Of Thwarted Expectations, and Harmful Exploitation", *Internet Policy Review*, Vol. 9, No. 4 (21 December 2020).

person's exercise of rights can affect other people – and vice versa´.

Within this context, the framing of privacy as an individual right presents a clear fallacy, and that is that it places the burden of control onto the individual in the face of an increasingly complex and opaque data extraction and handling. How is it conceivable for someone to identify, let alone challenge, the existing extraction of personal data which are inherently intertwined with all aspects of our life? This is a crucial issue we must reconcile with.

As AI systems increasingly place allocative functions in our society (on whether we receive a loan, for example), it appears increasingly complex to conceive privacy merely as a set of individual rights that one is entitled to exercise. How exactly is an individual supposed to understand the workings of an algorithm especially when it operates on inferential data, which may not even be fully considered personal data in the first place?[5]

Over recent times, a lot of discussions have focused on transparency, and transparency is indeed a key principle of the GDPR (General Data Protection Regulation). Transparency of an algorithm may be instrumental to understand where the bias may be coming from, and this is clearly present in recent regulatory developments. There are, however, serious issues with transparency.

First, achieving a balance between regulatory compliance, stakeholder interests and commercially sensitive information may be challenging. Second, transparency encompasses many different issues and aspects, with no one-size-fits-all model. This is because transparency is subjective and depends on existing knowledge, not least the willingness, available time and capacity to process new information and investigate a specific issue.

The transparency of an artifact arguably lies in the context of where it operates. For example, an AI tool used in a medical setting requires different degrees of transparency based on who operates it. Precision medicine software will need to provide a different set of information criteria to a patient as opposed to a medic. The patient is usually not medically trained and likely to be focused on whether it cured or helped with their problem, whereas the medic will need to take a more critical and expert stance, accepting liability for both the raw output and their operation and/or interpretation of the software.

[5] Sandra Wachter and Ben Mittelstadt, "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI", *Columbia Business Law Review*, No. 2 (2019): p. 494–620, available at: https://doi.org/10.7916/cblr.v2019i2.3424 (last accessed: 2o February 2022).

There could be a technical way to alleviate the first issue (achieving a balance between regulatory compliance, stakeholder interests and commercially sensitive information). That might be through multi-party computation and other cryptographic solutions can be used to share the algorithms in, for example, litigation whilst refraining from sharing it.

The second issue (the varying and subjective nature of transparency) remains a clear obstacle to exercise of an individual right. Namely, how is the individual supported by transparency? Clearly, an AI product is too complex for transparency to be viewed like a list of ingredients on a food label. What is needed, for that transparency to be really meaningful and enable the exercise of a right? This is the key question that we need to deal with; otherwise low quality 'transparency' is going to harm those who need transparency the most as they are at the sharp end of algorithmic discrimination.

Solove suggests adding a public dimension to privacy without eliminating the individual rights angle – the point is that we have to recognize its limits. The risk is that individuals penalized by AI systems could be blamed for not having enquired about the model themselves, which is plainly unfair in the current opaque scenario where these systems have so often been deployed and developed.

A Public Dimension of Privacy

Adding a collective and public dimension to privacy means two things:

First, it means that more efforts have to be placed in ex-ante measures. The Canadian Algorithmic Impact Assessment (AIA) for public sector algorithms is a good example, and so is the recent AIA piloted by the Ada Lovelace Institute for the UK's National Health Service (NHS). Ex-ante measures were once again the champions from a privacy standpoint in two recent rulings. In the first ruling, Uber drivers from the United Kingdom and Portugal asked for access to their personal data and the right to transparency of algorithmic management. The Amsterdam District Court ruled in favour of disclosing some of the personal data used in automated decision-making and profiling of drivers. In the second judgment, the same court ordered Uber to reinstall five drivers dismissed by the company's algorithmic system. It was one of the first cases that used Article 22 of the GDPR, which provides rights of protection against unfair automated decision making.

In both cases the ex-ante assessment of the algorithm was crucial to temper the risks of unfair treatment – in this vein, Frank Pasquale argues that a licensing agency should be set up to ensure only algorithms that have undergone due process

should be allowed to enter the market. The current draft EU AI Act proposes self assessments for high risk AI, except for facial recognition technology (FRT), following the product safety legislation already adopted in the EU through the CE marks.

"A licensure regime for data and the AI it powers would enable citizens to democratically shape data's scope and proper use, rather than resigning ourselves to being increasingly influenced and shaped by forces beyond our control", argues Pasquale.[6]

A new public dimension of privacy demands that we adopt a less static approach to privacy. Privacy as it is enshrined in legislation right now caters for privacy harms as they are now – in a sense, it champions the status quo. But the status quo is untenable in the age of AI and algorithmic decision making. I believe that a formulation of privacy as an ´ideal´ may open the door to sustainable AI. If we can reframe privacy more as the enhancing of personal agency and autonomy and less as merely an individual right to be demanded, then the protection of the data is more clearly a means to safeguard the people behind it and their agency.If this approach works out, we should be able to conceptualise what it means for our personal information to have a true collective value, and what controls, enforcement and limitations we put around such valuable assets. By partly releasing the individual from the responsibility over her/his own privacy and putting the onus back onto the organizations, we would be taking the important first step. With data already being viewed as a commodity, it will clearly be a major battle to bring in a public dimension of privacy, but it is perhaps our best bet if we are to build full public confidence and participation in the benefits of AI.

[6] Frank Pasquale, "Licensure as Data Governance - Moving toward an industrial policy for artificial intelligence", *Knight Columbia*, 28 September 2021, available at: https://knightcolumbia.org/content/licensure-as-data-governance (last accessed: 20 February 2022).

INTERNATIONAL COOPERATION ON ARTIFICIAL INTELLIGENCE- A PROBLEM STATEMENT

The following contribution sheds light on the norms at stake when it comes to international cooperation on artificial intelligence (AI). International cooperation on AI involves two conceptual pillars: international regulation and global diffusion of AI-powered technologies. While the international regulatory discourse is concerned with establishing frameworks for minimizing and mitigating negative consequences of AI development and deployment, the global diffusion of AI technologies presents global policy makers with an ethical dilemma. AI represents an important tool to realize social development, but also an instrument that can amplify pre-existing norm violations. Based on an analysis of the moral dilemmas posed by international cooperation on AI, the authors advocate focusing on the avoidance of complicity in human rights violations while maintaining technology transfer of AI solutions critical for human development.

Alexander Kriebitz* & Christoph Lütge**

TPQ
Winter 2021/22

* Alexander Kriebitz: Postdoctoral Researcher at Technical University of Munich.
** Christoph Lütge: Chair of the Peter Löscher Chair of Business Ethics and Director of the Institute of Ethics in Artificial Intelligence.

International cooperation on artificial intelligence (AI) has been hailed by international organizations, multinationals developing and deploying AI, universities and regulatory bodies. Moreover, international AI cooperation encompasses two key facets, namely, the international regulation of AI and its global diffusion.[1] In response to intensifying collaboration in both areas, human rights scholars and policy makers have voiced concerns over AI collaboration with states accused of violating international norms.[2] Based on human rights and national interest considerations, foreign policy makers have advocated measures to reduce international cooperation on AI, particularly with countries regarded as authoritarian.[3] Owing to this dissonance between emerging AI cooperation and efforts to curb the global diffusion of AI, the following contribution sheds light on the norms at stake when it comes to international cooperation on AI. The paper articulates the ethical problem statement of international cooperation on AI pertaining to the two following questions:

- What aspects do organizations need to consider when interacting commercially with actors outside of the OECD?

- Is it legitimate to engage in technology transfer on AI with authoritarian regimes?

The remainder of the publication is structured as follows: The paper begins with a review of the existing discourse on international technology cooperation to delineate potential areas of norm collisions that might be relevant for AI. In a second step, the contribution concentrates on the impact analysis of AI on the discussed norms. On this basis, the authors examine the prospective role of proportionality as a principle to reconcile conflicting norms and interests.[4]

Normative Concepts in International Affairs

In the following contribution, we understand "normative" as a general property pertaining to morally desirable aims, situations, or actions.[5] Normative considerations

[1] World Bank. *Better Data for Doing Good: Responsible Use of Big Data and Artificial Intelligence*, OECD/LEGAL/0449, (2018).

[2] A. Polyakova and C. Meserole, *Exporting Digital Authoritarianism: The Russian and Chinese Models. Policy Brief, Democracy and Disorder Series* (Washington, DC: Brookings, 2019): p. 1-22; G. King, J. Pan and M. E. Roberts, "How Censorship in China Allows Government Criticism but Silences Collective Expression," *American Political Science Review*, Vol. 107, No. 2 (Month 2013): p. 326-343.

[3] H.R. 3426 (IH) - *Democracy Technology Partnership Act*.

[4] S. S. ÓhÉigeartaigh et al., "Overcoming Barriers to Cross-Cultural Cooperation in AI Ethics and Governance," *Philosophy & Technology*, Vol. 33, No. 4 (Month 2020): p. 571-593.

[5] C. Lütge, *Order Ethics or Moral Surplus: What Holds a Society Together?* (Lexington Books, 2015); C. Lütge and M. Uhl, *Business Ethics: An Economically Informed Perspective* (USA: Oxford University Press, 2021).

concern various aspects of AI cooperation beyond AI regulation, namely, the diffusion of AI solutions on a global level by investment, finance, trade, or development aid. These modes of cooperation are international in nature and hence fall under the domain of international ethics. Thus, the UN Charter commonly understood as the gravitational center of this normative space develops programmatic goals the United Nations ought to realize in Article I.[6] Moral desiderata of international affairs include international peace (UN Charter, Art. 1[1]); the solution of "international problems of an economic, social, cultural, or humanitarian character" (UN Charter, Art. 1[3]), and "respect for human rights" (UN Charter, Art. 1[3]). International treaties have further specified the aims articulated in Article I, including the Universal Declaration of Human Rights (1948), the International Covenant of Political and Civic Rights (1967), and the International Covenant of Economic, Social and Cultural Rights (1967). International AI cooperation is situated within this comprehensive normative corpus.

"International cooperation deals with the misuse of technologies by engaging against cybercrime or preventing distribution of weapons of mass destruction to terrorist organizations."

The Normative Debate on Technology Cooperation

Technologies matter for the (non-)realization of moral aims in international affairs, creating significant repercussions for individuals, collective organizations, or the planet as a whole.[7] Besides, technologies produce intended and unintended consequences.[8] Unintended consequences may be disasters caused by technical error, human miscalculations, or ignorance of long-term consequences – including "ultra-hazardous activities"; intended harm pertains to the deliberate use of technologies at the expense of the rights and interests of others.[9]

Consequently, international cooperation matters for harnessing technology in a way

[6] A. Verdross, "General International Law and the United Nations Charter," *International Affairs* (Royal Institute of International Affairs 1944-1954): p. 342-348; S. Chesterman, I. Johnstone and D. Malone, *Law and Practice of the United Nations: Documents and Commentary* (Oxford: Oxford University Press, 2016).

[7] H. Jonas, *Toward a Philosophy of Technology* (Hastings Center Report, 1979): p. 34-43.

[8] R. N. Osborn and D. H. Jackson, "Leaders, Riverboat Gamblers, or Purposeful Unintended Consequences in the Management of Complex, Dangerous Technologies," *Academy of Management Journal*, Vol. 31, No. 4 (1988), p. 924-947.

[9] A. E. Boyle, "Globalising Environmental Liability: The Interplay of National and International Law," *Journal of Environmental Law*, Vol. 17, No. 1 (2005): p. 3-26.

that does not threaten but enhances major aims in international affairs. This importance expresses itself in various channels and formats of collaboration: International frameworks and procedures are intended to mitigate risks of technologies, when human and technical error might cause immense damage.[10] In addition, interdependencies render international regulation necessary when it comes to basic traffic rules, the mitigation of natural catastrophes by early warning systems, or collaboration in space. Likewise, international regulation aims at incentivizing technological progress through intellectual property rights (WIPO) or frameworks enabling international competition. International cooperation deals with the misuse of technologies by engaging against cybercrime or preventing distribution of weapons of mass destruction to terrorist organizations.[11] Thus, international cooperation involves norm enforcement and creation. Based on the latter pillar, the international community has formulated legal provisions against biological weapons, human cloning, or lasers to blind human beings.[12]

The Necessity of International Technology Transfer

Technology transfer depicts a quintessential aspect of international technologies cooperation as it pertains to the global diffusion of technologies, skills, and knowledge.[13] A strong argument for international technology transfer emerges from the UN Charter outlining the purpose of international cooperation in terms of "solving international problems of an economic, social, cultural, or humanitarian character*"* (UN Charter Art. 1[3]).

Economic imbalances between developing and developed nations depict one of the most pressing issues in contemporary international affairs, manifesting in different levels of life expectations, education, growth rates and literacy or access to medical treatment and food.[14] The normative aim to realize a fair distribution of global capabilities, resources, and wealth manifests itself in the notion of a "right to development." The Declaration on the Right to Development adopted by the UN

[10] S. Miller, *Dual Use Science and Technology, Ethics and Weapons of Mass Destruction* (Dordrecht: Springer, 2018): p. 2018.

[11] G. Calcara, "The Role of INTERPOL and Europol in the Fight Against Cybercrime, with Reference to the Sexual Exploitation of Children Online and Child Pornography," *Masaryk University Journal of Law and Technology*, Vol. 7, No. 1 (2013): p. 19-33.

[12] M. H. Arsanjani, "Negotiating the UN Declaration on Human Cloning," *American Journal of International Law*, Vol. 100, No. 1 (2006): p. 164-179; J. Goldblat, "The Biological Weapons Convention: An Overview," *International Review of the Red Cross (1961-1997)*, Vol. 37, No. 318 (1997): p. 251-265; J. H. McCall Jr, "Blinded by the Light: International Law and the Legality of Anti-Optic Laser Weapons," *Cornell Int'l LJ*, Vol. 30, No. 1 (1997).

[13] J. Gottwald, L. F. Buc and W. Leal Filho, "Technology Transfer," in S. O. Idowu, N. Capaldi, L. Zu, and A. D. Gupta (eds.), *Encyclopedia of Corporate Social Responsibility* (Berlin, Heidelberg: Springer, 2013). https://doi.org/10.1007/978-3-642-28036-8_673

[14] A. Sengupta, "On the Theory and Practice of the Right to Development," *Human Rights Quarterly,* Vol. 24, No. 4 (2002): p. 837-889.

General Assembly in 1986 refers here to "equality of opportunity for all in their access to basic resources, education, health services, food, housing, employment and the fair distribution of income" (Art. VIII [1]).[15] The postulated right encapsulates the idea that "nations have a duty to contribute to the assimilation of living standards internationally."[16] Thus, literature suggests that basic innovations play a critical role in realizing similar living standards in the international context and as an instrument to realize long-term growth.[17] Moreover, the notion of inequalities based on access to technologies has been articulated in the term "global digital divide," implying that access to certain technologies characterizes different living standards.[18] Moreover, the International Covenant on Economic, Social and Cultural Rights defines rights (including rights to health or food) as concrete aims of national and international efforts. Denying developing countries access to technologies in critical areas such as technologies for vaccine distribution would constitute human rights violations.[19]

"The right to development and human rights in their social and economic dimensions correspond to the interest of the global majority in establishing global minimum standards. Here, decision makers should refrain from unilateral measures enhancing preexisting technological divides."

The Discontents of International Technological Cooperation and Technology Transfer

In spite of the prospective impact of technology on social development, international observers have voiced concerns over technology transfer on moral grounds. [20] This applies specifically to the view that technologies might threaten the realization

[15] I. D. Bunn, "The Right to Development: Implications for International Economic Law," *Am. U. Int'l L. Rev*., Vol. 15, No. 1425 (1999).

[16] A. Sengupta, "Implementing the Right to Development," *International Law and Sustainable Development* (Brill Nijhoff, 2004): p. 341-377.

[17] R. M. Solow, "Technical Change and the Aggregate Production Function," *The Review of Economics and Statistics*, (1957): p. 312-320; R. M. Solow, "A Contribution to the Theory of Economic Growth," *The Quarterly Journal of Economics*, Vol. 70, No.1 (1956): p. 65-94.

[18] W. Chen and B. Wellman, "The Global Digital Divide–Within and Between Countries," *IT & Society*, Vol. 1, No. 7 (2004), p. 39-45.

[19] U.N. CHR, General Comment 31, "The Nature of the General Legal Obligation Imposed on States Parties to the Covenant," Article 2, (29 March 2004), CCPR/C/74/CRP.4/Rev.6.

[20] D. Scissors and S. Bucci, "China Cyber Threat: Huawei and American Policy Toward Chinese Companies," *The Heritage Foundation* (2012): p. 3761; Polyakova and Messerole (2019).

of international peace, human rights, or national interest. This belief originates in historical perceptions: Nuclear weapons, cyberwars, and militaries have reconfigured international power structures, and technologies might empower actors in international affairs with an agenda adverse to human rights or international peace. This concerns not only diffusion of lethal weapons, but also the spread of dual-use items and fundamental science. Likewise, technologies can bear major ramifications for situations inside of authoritarian regimes when directed against individual rights. Consequently, literature in the business and human rights discourse embarked on exploring the implications of the potential misuse of technology transfer by authoritarian leaders.[21] The Guiding Principles on Business and Human Rights articulates here the view that business operations "should address adverse human rights impacts with which they are involved." Efforts to limit technology transfer originate in different ideas: preventing direct complicity in human rights/humanitarian law violations and containing geopolitical adversaries.

International Cooperation on AI as a Moral Dilemma

The discourse on technology transfer confronts decision makers with conflicting implications: Technologies require a set of global minimum standards, their diffusion is vital for the convergence of developing nations, and they might amplify preexisting tendencies toward norm violations. Consequently, the remainder of the paper is focused on the prospective role of AI within this context.

The Need for International Regulation of AI

AI combines large amounts of data with intelligent algorithms, allowing the software to learn automatically from given data input.[22] A further communality of AI solutions is based on their capability to imitate intelligent human behavior.[23] Consequently, literature defines AI as a "family of technologies," implying that AI-based human rights violations may have different causes and origins.[24] Specifically, AI can trigger unintended consequences—for example, biases may trigger discrimination, or organizations and agents can use AI with the explicit intention to create damage.[25]

[21] UN Guiding Principles on Business and Human Rights Principle 11.

[22] Compare: EU AI Act; R. S. Michalski, J. G. Carbonell and T. M. Mitchell (eds.), *Machine Learning: An Artificial Intelligence Approach* (Springer Science & Business Media, 2013).

[23] A. Kaplan and M. Haenlein, "Siri, Siri, In My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence", *Business Horizons,* Vol. 62, No. 1 (2019): p. 15-25;
C. Bartneck, C. Lütge, A. Wagner and S. Welsh, *An Introduction to Ethics in Robotics and AI* (Springer Nature, 2021) p. 117.

[24] A. Kriebitz and C. Lütge, "Artificial Intelligence and Human Rights: A Business Ethical Assessment," *Business and Human Rights Journal*, Vol. 5, No. 1 (2020): p. 84-104; Bartneck, Lütge, Wagner, and Welsh (2021), p. 117.

[25] A. Lambrecht and C. Tucker, "Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads," *Management Science*, Vol. 65, No. 7 (2019): p. 2966-2981.

Consequently, international cooperation needs to cover both aspects. The black-box character of AI and questions linked to the human oversight–responsibility nexus necessitate common guidelines and codes of conduct as minimum standards for AI development and deployment. Interdependency constitutes a further variable in the equation, determining the need for international AI cooperation. This applies in specific to the international data transfer. The purpose of such cooperation would be to render AI technologies safer and better suited for realizing moral aims by establishing principles for a global AI ecosystem and outlawing misuse of AI, such as subliminal techniques, manipulative technologies, or certain supervision technologies.[26] A precondition of this regulatory discourse is to comprehend exactly how AI affects collective aims.

AI as an Enabler of Collective Aims

Preceding literature illustrates that artificial intelligence is conducive to the realization of human rights, social development, and economic growth.[27] Given its economic relevance, studies have warned that an imbalanced distribution of AI could exacerbate preexisting inequalities on a global level. Future differences in living standards could therefore express themselves in how societies adapt and use artificial intelligence. Therefore, it is plausible to connect the notion of "right to development" with policies incentivizing a global diffusion of vital AI technologies that safeguard a global minimum of AI deployment.[28] Moreover, AI solutions often address particular aspects relevant to human rights realization. This applies specifically to the right to health, where researchers highlight a range of technologies conducive to global health.[29] For example, AI solutions could help to detect tuberculosis or cervical cancer.[30] Furthermore, autonomous driving depicts a basic technology enhancing passenger and traffic safety.[31] Consequently, collaboration on AI is of specific importance for resource-poor countries.[32]

[26] World Health Organization*, Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*, (2021).

[27] Y. He, "The Importance of Artificial Intelligence to Economic Growth," *Korea Journal of Artificial Intelligence*, Vol. 7, No. 1 (2019): p. 17-22; J. Bughin et al., *Artificial Intelligence the Next Digital Frontier?* (McKinsey Global Institute, 2017); PwC, *The Macroeconomic Impact of Artificial Intelligence* (2018).

[28] A. Korinek and J. E. Stiglitz, *Artificial Intelligence, Globalization, and Strategies for Economic Development*, (No. w28453), (National Bureau of Economic Research, 2021).

[29] A. Ismail and N. Kumar, "AI in Global Health: The View from the Front Lines," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (May 2021): p. 1-21.

[30] N. Schwalbe and B. Wahl, "Artificial Intelligence and the Future of Global Health," *The Lancet*, Vol. 395, No. 10236 (2020): p. 1579-1586.

[31] C. Lütge, "The German Ethics Code for Automated and Connected Driving," *Philosophy & Technology*, Vol. 30, No. 4 (2017): p. 547-558.

[32] B. Wahl et al., "Artificial Intelligence (AI) and Global Health: How Can AI Contribute to Health in Resource-Poor Settings?," *BMJ Global Health*, Vol. 3, No. 4 (2018), e000798.

AI as a Threat to Collective Aims

As mentioned above, AI represents a comprehensive set of technologies, with certain AI solutions contravening human rights and international law. AI-driven supervision technologies, such as certain manipulative AI technologies or the use of AI-powered lie detectors or social credit ratings, have repeatedly sparked ethical concerns due to their invasiveness of privacy and related to freedom of expression.[33] A further area of concern is the use of neutral technologies (such as face recognition) for realizing controversial aims. This applies to settings in which powerful actors have interests in using AI to enhance their control over others.[34] In particular, Western NGOs and scholars have scrutinized a wide range of countries, including Angola, Ethiopia, China, Russia, and Venezuela concerning developing, using, or exporting AI-driven technologies in ways that infringe upon human rights.[35] Nevertheless, similar incentive structures might prevail beyond countries classified as authoritarian, as actors in democracies could be tempted to (mis-)use similar technologies in moments of crisis.[36] Similar arguments have been raised in the military context of AI: AI might not only be used for upgrading conventional weapons, but also to coordinate cyber-attacks or as robot soldiers. Scholars have warned that AI-based robots would have difficulties distinguishing between civilians and combatants, enhancing risks of breaches of the Geneva Convention.[37] Consequently, uncontrolled AI diffusion could have adverse impacts on human rights and humanitarian international law, owing to its nature of standardizing processes and reducing human oversight.

A Sketch for Ethical AI Cooperation

Comparing the technology transfer discourse and the debate on AI ethics reveals the moral dilemma decision makers face when dealing with international cooperation and technology transfer. Thus, AI raises the stakes in the pre-existing collision between different norms and interests of international cooperation. In the following, we sketch a set of preliminary suggestions derived from the parallelism of technology transfer and AI ethics discourses.

---

[33] E. Donahoe and M. M. Metzger, "Artificial Intelligence and Human Rights," *Journal of Democracy*, Vol. 30, No. 2 (2019): p. 115-126.

[34] A. Kriebitz and R. Max, "The Xinjiang Case and Its Implications from a Business Ethics Perspective," *Human Rights Review,* Vol. 21, No. 3 (2020): p. 243-265.

[35] J. Millward and D. Peterson, *China's System of Oppression in Xinjiang: How it Developed and How to Curb it* (Brookings Institution, 2020); A. Polyakova and C. Meserole, *Exporting Digital Authoritarianism: The Russian and Chinese models* - Policy Brief, Democracy and Disorder Series (Washington, DC: Brookings, 2019): p. 1-22.

[36] P. Molnar, *Robots and Refugees: The Human Rights Impacts of Artificial Intelligence and Automated Decision-Making in Migration – Research Handbook on International Migration and Digital Technology* (Edward Elgar Publishing, 2021).

[37] U. Pagallo, "Robots of Just War: A Legal Perspective," *Philosophy & Technology*, Vol. 24, No. 3 (2011): p. 307-323.

Mitigating Unintended Consequences of AI

Irrespective of the particular interests of AI developers or users, all sides share the aim of mitigating unintended consequences of AI use. This applies, for instance, to the regulation of biases causing discrimination or to unsafe AI solutions threatening human life in areas such as health or autonomous driving. The search for global minimum standards constitutes a further prospective area of global AI regulation. Prospective candidates of international condemnation and prohibition could be subliminal AI techniques, manipulative AI solutions, or AI-powered lie detectors, but also the more general question of (biometric) data input. Furthermore, international regulation is relevant for AI use in warfare, but supervision technologies are also of critical importance for realizing existent codifications on human rights or the *jus in bello*.

Identifying Normative Boundaries

Normative limitations pertain to areas where AI cooperation or detrimental to *moral desiderata* in international affairs. The right to development and human rights in their social and economic dimensions correspond to the interest of the global majority in establishing global minimum standards. Here, decision makers should refrain from unilateral measures enhancing preexisting technological divides. Likewise, uncontrolled technology transfer can exacerbate preexisting human rights violations. Following the UN Guiding Principles (GPs), organizations ought not to be complicit in any human rights violations, implying that the proliferation of certain AI technologies to authoritarian regimes – such as AI-based lie detectors or AI solutions involving biometric data – would violate the UN GPs.

Involving Proportionality

Both implications are likely to collide in practice, especially when human rights-based restrictions on AI diffusion drive global inequalities. Hence, a balanced assessment requires minding the interest of individuals living in authoritarian regimes. As an approximation, we formulate the following premises as variables in the equation of limiting or enhancing technology transfer.

1. *Developing countries* (even the less democratic ones) require technological support in terms of AI solutions for closing preexisting gaps. Political concessions such as guarantees not to use AI solutions in specific domains could be realized within this collaborative format.

2. *Certain political structures* (with a low degree of checks and balances and low regard for human rights) and the existence of preexisting conflicts

present determine the likelihood of AI development and deployment at the expense of human rights and humanitarian law.

3. Specific AI solutions are *closely related to aims in international affairs*. This includes AI solutions to climate change (smart energy solutions) and those conducive to the resolution of a health crisis (AI-powered vaccine development). These AI solutions should be prioritized within global AI cooperation.

4. Areas of AI use that are generally deemed as *conflicting with human rights* include certain supervision technologies or the use of biometric data. Regulation on such technologies presents an urgent matter for international discourse.

These points might enable a cost/benefit assessment of AI cooperation from a moral perspective. However, international regulation and global diffusion of AI hinge on each other, as stalling international regulation might delay technology diffusion. Likewise, limiting technology transfer might hinder international alignment in a setting involving not only traditional powers such as the U.S., EU, and Russia, but also emerging leaders in particular fields of AI such as China, Turkey, and South Korea.

Conclusion

The question of how to engage in international cooperation on AI is an intricate issue. International AI cooperation itself is a complex topic involving two interdependent pillars, regulation and diffusion of AI. The preceding analysis indicates that norms are limiting but also guiding potential areas of AI cooperation. The configuration of AI as a basic innovation that might create a series of unintended consequences requires a higher degree of collaboration than other technologies. This requirement is exacerbated by interdependences and the urgency of global standards on critical AI solutions. AI diffusion is even more complex, due to the nature of AI as a powerful tool for development and its role in human rights violations. Policy makers need to include both aspects, representing different types of rights and interests in their calculations. For solving conflicts between both pillars, it is therefore important to distinguish not only between authoritarian and democratic nations, but also between developing and developed nations.

While the principle of proportionality presents a starting point for navigating international AI cooperation, more research needs to be undertaken to gain a more precise picture of further potential trade-offs in the global diffusion of AI.

PROACTIVE MANAGEMENT OF CREATIVE DESTRUCTION: TRANSPARENCY AS A FOUNDATIONAL ELEMENT

AI is very much still in its infancy. The path to beneficial and true development lies with transparency and regulation for both governments and companies. There is a wide-spread mistrust of "artificial intelligence" as a technology, and the companies that are developing it, across many stakeholders. There is a similar mistrust of governments' roles in or ability to regulate tech companies or to govern their own use of these same technologies. Regulation, even though is the key, should be implemented rather carefully so that development is not hindered but is steered in the right direction.

Stacey H. King*

TPQ
Winter 2021/22

* Stacey H. King, Visiting Policy Fellow – Oxford Internet Institute.

When entering into any kind of multi-party business relationship, transparency among all parties is critical to the success of the outcome. For example, if you want to build a house, making sure the architect, engineer, contractor, city planner, landscape designers, and builders are all onboard with your anticipated needs and future potential use of the house is necessary. Without such transparency, foundations can fail, walls can crack, plumbing can leak, extensions are rejected, and, in some cases, neighbors may complain and prevent plans from moving forward. The same need for transparency and honest dealings is true for the regulation of artificial intelligence (AI).

AI is very much still in its infancy. The AI being built today – narrow AI (ANI), or AI that is taught to manage a particular task – is the foundational building block for the AI of tomorrow – artificial general intelligence (AGI), or AI that is able to replicate broader human thought processes across multiple tasks. Exactly how both narrow and general AI will impact global economies and organizational structures in the short and long-term is still unknown,[1] but recent developments provide enough insight to know that AI will bring not just a gale, but a tsunami of Schumpeter's creative destruction.[2] Indeed, we may currently be standing at the shore of Schumpeter's gale, watching the water pull away toward sea.

Now is the critical time, while this technology is still in its infancy, for all stakeholders – governments, industry, academia, and civil society – to be more forthcoming and transparent about the knowns and unknowns of AI and its potential impact, to collaborate on ways to support responsible development and use of the same, and to prepare for the inevitable changes to come. But to do so, all stakeholders must embrace a requisite level of transparency, otherwise the walls of this house will inevitably bend under pressure or fail to get off the ground.

Creative Destruction

Just as businesses must organize around and execute on strategies to manage change and protect their relevance and position accordingly,[3] so too should governments and society. AI differs from previous disruptive innovations to both market and non-market approaches. Automation itself is nothing new. Humans have automated tasks for centuries and, while initially disruptive, have modified or introduced new ways or organizing our societies accordingly.[4] More traditional forms of automation sought

[1] Nicholas Crafts, "Artificial Intelligence as a General-purpose Technology: An Historical Perspective," *Oxford Review of Economic Policy*, Vol. 37, No. 3 (2021): p. 521-36.

[2] J. A. Schumpeter, (1943), *Capitalism, Socialism, and Democracy* (London: Routledge Classics, 2010 ed.).

[3] C. O'Reilly and M. Tushman, "Lead and Disrupt: How to Solve the Innovator's Dilemma," (Stanford: Stanford Business Books, 2016).

[4] C. B. Frey, & M. A. Osborne, "The Future of Employment: How Suceptible Are Jobs to Computerisation?," *Technological Forecasting & Social Change*, (2017): p. 254-280.

to replicate "physical" processes (reproducing works, shipping products, drilling holes in steel, amplifying your voice, or searching through books in a library). And in response society amended or developed structures, regulations, industry standards, and ethical norms for managing the automation of physical processes.

> *"Artificial intelligence as used today is adding to this mix in ways that are not always perceived as beneficial - from surveillance to deep fakes to harmful bias and discrimination."*

Artificial intelligence, however, is a new type of automation with an important distinction. AI does not seek to replace another human physical process, per se, but rather to replicate part of the human "thought" process or the decision-making behind the doing of something.[5] This is a fundamentally different form of automation – it challenges the very notion of what humans believe distinguishes them as a species – and our societies do not have as many mechanisms in place for managing the fallout from this type of technology. As a result, transparency around the foundational parts of this technology (ANI) and how it is developing (AGI) by those closest to it – the architects and engineers of this house – is more critical than ever.

The market for AI in relation to widespread use across industries, products, and services is still early.[6] Nevertheless, wider concerns about the potential impact of AI have stretched from claims that machines will replace all jobs and, eventually, humans,[7] to claims that machines will replace jobs but lead to an era of creativity and freedom.[8] In both cases the supposed outcome is based on conjecture and myth, but they are important myths to recognize as they define the fears and hopes that many stakeholders bring to the table and influence how we approach the development and regulation of AI as a society.

[5] "The capacity of computers or other machines to exhibit or simulate intelligent behaviour," IEEE, *Ethically Aligned Design: First Edition Glossary*, p. 8. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead1e_glossary.pdf [visited Feb. 29, 2022].

[6] D. Nathan and N. Ahmed, "Technological Change and Employment: Creative Destruction," *The Indian Journal of Labour Economics*, (2018): p. 281-298.

[7] P. Holley, "Elon Musk's nightmarish warning: AI could become 'an immortal dictator from which we would never escape," 6 April 2018, Retrieved from *The Washington Post*: https://www.washingtonpost.com/news/innovations/wp/2018/04/06/elon-musks-nightmarish-warning-ai-could-become-an-immortal-dictator-from-which-we-would-never-escape/?utm_term=.83898a54f99f

[8] S. V. Osborn, "The End of Meaningless Jobs Will Unleash the World's Creativity," Retrieved from SingularityHub, 23 August 2016: https://singularityhub.com/2016/08/23/the-end-of-meaningless-jobs-will-unleash-the-worlds-creativity/#sm.0001enmry1fb5dfsxfw17285ra9ll

AI will have a wide array of impacts that we cannot predict.[9] We do know from previous general-purpose technologies (steam, railroad, electricity, Internet) that societies and economies are rarely prepared to deal with the wide-ranging impacts such technologies bring.[10] The fallout does not impact one industry or market or regulation.[11] It is this wide-scale impact that we, as a society, need to address now. If we know AI is going to have such broad implications, how do we prepare for the negative impacts of a radical shift in the status quo, while simultaneously encouraging innovation and market growth? And why is transparency critical to the same?

State of Play Today

Our global society is already in a period of upheaval, in part arguably a long-tail impact of the introduction of the Internet.[12] The "technology revolution" introduced a myriad of opportunities, economic advantages, and the ability to connect on a global basis, but it was also disruptive to existing societal systems and governance, led to incursions on individuals' private lives, and has amplified the reach of those who do not have societies' best interests at heart. The latter has led to a multitude of finger pointing and, in many cases, the refusal of companies to participate transparently in the debate over unintended consequences and how to remedy them.

Artificial intelligence as used today is adding to this mix in ways that are not always perceived as beneficial - from surveillance to deep fakes to harmful bias and discrimination.[13] In addition, the division of wealth (personal, corporate, and national) is becoming more extreme, market skills are changing, and industrial society norms are being questioned. Finally, communities are becoming more and more polarized, and the environment has become a critical threat to livelihood and economies. Individuals feel less in control of what is happening around them, the infrastructure that is supposed to support them, security for the future, and more and more mistrust of technology itself. And in the midst of all of this, we are asking

[9] P. Aghion, U. Akcigit and P. Howitt, "What Do We Learn From Schumpeterian Growth Theory?," Working Paper 18824. Cambridge: National Bureau of Economic Research (2013).

[10] C. Schubert, "How to Evaluate Creative Destruction: Reconstructing Schumpeter's Approach," *Cambridge Journal of Economics*, (2013): p. 227-250.

[11] P. Aghion, U. Akcigit and P. Howitt, (2013).

[12] Rostam J. Neuwirth, "The 'letter' and the 'spirit' of Comparative Law in the Time of 'artificial Intelligence' and Other Oxymora," *Canterbury Law Review*, Vol. 26, (November 2020): p. 1-31.

[13] E.g., "AI is Sending People to Jail – and Getting It Wrong," *MIT Technology Review*, 21 January 2019, https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/ [visited on 18 February 2022]; "This is What a Deepfake Voice Clone Used in a Failed Fraud Attempt Sounds Like," *The Verge*, 27 July 2020, https://www.theverge.com/2020/7/27/21339898/deepfake-audio-voice-clone-scam-attempt-nisos [visited on 18 February 2022]; "A.I. Bias Caused 80 percent of Black Mortgage Applicants to Be Denied," *Forbes*, 2 September 2021. https://www.forbes.com/sites/korihale/2021/09/02/ai-bias-caused-80-of-black-mortgage-applicants-to-be-denied/?sh=2f9b093136fe [visited on 18 February 2022]. "Exploiting AI: How Cybercriminals Misues and Abuse AI and ML," 19 November 2020. https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/exploiting-ai-how-cybercriminals-misuse-abuse-ai-and-ml [visited 19 February 2022].

our larger societies to trust the companies building systems to eventually replicate human thought process and make decisions, even though little is known about the foundations on which this systems are built and who is training them to make the decision.[14] To develop global trust that these systems are being built for human good, and not just to drive revenue or reinforce a government's power, relevant stakeholders need to be more transparent and provide safety-nets and compliance mechanisms for managing negative impacts. Transparency does not mean that stakeholders need to disclose every business plan or trade secret. Transparency does require stakeholders to come to the table in an open and honest manner, and to negotiate the issues that arise accordingly.

> *"As algorithms trained by different developers to manage specific tasks become integrated into larger systems, and as these systems become more independent from human intervention in decision-making, questions around jurisdictional oversight will become of greater concern."*

Why are these factors we face today so important to understand in creating a strategy for the promotion of cooperative outcomes in governance for the future? Because it is precisely these issues that lay under the surface of every stakeholder critical to the systems that help us navigate through the changes to come.

Mistrust of AI, Tech CoDampanies, and Government

There is a wide-spread mistrust of "artificial intelligence" as a technology, and the companies that are developing it, across many stakeholders.[15] At the same time, most stakeholders understand that AI is not going to disappear nor is it something we necessarily want to stop. What impacts stakeholder trust is the relatively recent, positive, and hopeful attitude toward technology that was found in the late 20th and early 21st Century, but which has been replaced with a general mistrust or concern around the same today. This leads to a perception that tech companies are concerned more about profit over people and society, are secretive and therefore suspect, and use "self-regulation" as a means for anti-competitive or anti-consumer protection behavior.[16]

[14] Jenny Bunn, "Working in Contexts for Which Transparency Is Important: A Recordkeeping View of Explainable Artificial Intelligence (XAI)," *Records Management Journal*, Vol. 30, No. 2 (2020): p. 143-153.

[15] "Trust in Artificial Intelligence: a Five Country Study," KPMG/University of Queensland, Australia (March 2021). https://assets.kpmg/content/dam/kpmg/au/pdf/2021/trust-in-ai-multiple-countries.pdf [visited on 9 February 2022].

[16] KPMG/University of Queensland, Australia (March 2021).

There is a similar mistrust of governments' roles in or ability to regulate tech companies or to govern their own use of these same technologies. On one hand, many are concerned that industry has captured regulators through lobbyists and market power. On the other, there is concern that governments will misuse the technology to exercise control over its citizens or to hold other nations hostage. Governments, like tech companies and industry, need to put covert nationalistic interests to the side and be more transparent about the pressures they face, collaborate on the larger societal needs, and create foundational structures of governance for all.[17]

National Strategies Focused on AI Fears

A result of this lack of trust is the development of national regulatory and economic strategies that often reflect fear mitigation as opposed to a holistic and informed approach to how we drive a beneficial development, deployment, and use of AI. In many cases, regulatory approaches are based on short-term thinking around a nascent technology. Overly strict regulations will limit our ability to develop this technology to its full potential or may result in nationalistic or inappropriate use or access by government. Adopting a wait-and-see/let the market settle itself out approach, however, may result in a lack of guardrails needed for the responsible development and use of this technology, and lead to similar misuse by stakeholders.

Conflict of Laws and Standards

Today over 700 initiatives across 60 nations are in some stage of development.[18] These include national and international regulatory strategies, industry standards, public consultations, and development of oversight boards. Almost 250 of these initiatives are national strategies and plans, each of which will reflect the foundational theories of governance reflected across the disparate jurisdictions: in other words, laws developed around theories of individual rights, market-based drivers, or security of the state over the individual. Just as we have seen the difficulties companies and governments have faced in regulating the Internet and world-wide-web given its global reach, so too will we face similar issues with AI.

As algorithms trained by different developers to manage specific tasks become integrated into larger systems, and as these systems become more independent from human intervention in decision-making, questions around jurisdictional oversight will become of greater concern. In addition, how companies will implement the volume of regulations – many targeting disparate pieces of a system (datasets vs.

[17] Steve Mills, Matthew Mendelsohn, et al., "Responsible AI Builds Trust in Government," *BCG*, January 2021, https://web-assets.bcg.com/98/15/18c0ec044bb7b9e2a372ea036371/bcg-responsible-ai-builds-trust-in-government-jan-2021-r.pdf [visited 2 February 2022].

[18] National AI Policies and Strategies, https://oecd.ai/en/dashboards [visited on 19 February 2022].

algorithms vs. use vs. data collection vs. access vs. documentation) – is another question. Without coordination across stakeholders, we risk either (a) regulatory interventions with compliance difficulties or (b) a race to develop an initial standard that requires global compliance, like the GDPR, but which may impact future development and use of the technology.

Academia and Industry: A Love-Hate Relationship

Academia and industry are also facing bottlenecks in driving collaboration. Industry is often hesitant to open itself to researchers or academic institutions as it may lead to a perceived lack of competitive advantage, criticism that could impact revenue or market price, or slow a process already in flight. Industry often requires researchers to work with strict guidelines when collaborating, which calls into question the objectivity and accuracy of resulting research and reports. For academic institutions and research that do not take corporate funding and cooperation for research, the results may be well intentioned recommendations for a path forward, but often are theoretical and disconnected from technical or organizational practices and implementation.[19] As a result, outcomes continue to result in silos that do not meet the needs for moving us forward in a meaningful way.

Self-Regulation

In many ways the rise of the Internet and the technologies that it drove biased us toward self-regulation as the most practical approach to early-stage innovation. But we are now experiencing the dark side of that self-regulation, which has forced us to question if self-regulation is still an appropriate path. Indeed, many of the companies,[20]governments, and individuals that benefited most from self-regulation have acknowledged that with AI, some form of regulation is necessary. The question over what types of regulatory guidelines around responsible development and use of technology the tech industry must implement is one that no private company will voluntarily put in place if it perceives it will result in a disadvantage in favor of a competitor. We must therefore carefully balance the important rights of a private company to act as a private company if we are to expect transparency from the same.

Private organizations are not non-profit organizations, nor are the tools they develop public utilities. In general, companies build products to meet a customer need (real or anticipated) and to earn revenue from the same. They do not generally develop

[19] Jessica Morley, Luciano Floridi, Libby Kinsey and Anat Elhalal, "From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices," *Science and Engineering Ethics*, Vol. 26, No. 4 (2019): p. 2141-168.

[20] See, e.g., "Microsoft President Brad Smith Calls for AI Regulation at Davos," *GeekWire*, 21 January 2020, https://www.geekwire.com/2020/microsoft-president-brad-smith-calls-ai-regulation-davos/ [visited on 18 February 2022].

products for nefarious purposes. But, as we have seen, there are unforeseen consequences that arise when we do not anticipate potential misuses or issues that arise in development processes or product use. Companies often approach development, particularly around technology, as a puzzle to solve. The question of whether it *should* be solved is one that, particularly in anticipation of AGI, becomes potentially more existential than with other technologies and certainly requires a different approach from "move fast, break things, and fix the bugs later." It serves as a fundamental shift in strategic thinking and one that will take time, resources, and organizational change to make a natural part of the development or business practice. As a result, there is a hesitancy to change, to be more transparent, or to close any door to a potential business opportunity in the future – known or unknown – which has surfaced as a lack of meaningful collaboration with other stakeholders. And this, perhaps, is one of the biggest potential roadblocks we must overcome.

Where Do We Go from Here?

We need to regenerate the willingness of stakeholders to work together towards a common ground and to be more transparent with one another to form a workable solution. By pro-actively organizing and working across stakeholders, we can help drive the development of the governance systems we will need both today and tomorrow. Industry and government need to be more transparent and forthright in where we are and where we are going and willing to address the question of *should* we proceed with something (even if it will meet a consumer or shareholder's needs), and academia and civil society need to be willing to develop realistic regulations that are implementable, do not slow innovation or create more problems in development, or redefine how a private company operates and governs.

So how do we do this when stakeholders have in many respects gone to their corners? First, education is a critical part of collaboration. Where the concentration of technical expertise lies increasingly within smaller and smaller groups, collaboration becomes an uneven playing field. Education and dialogue need to be at a level that is accessible to a wide-variety of stakeholders and not (a) purely academic, (b) understood only by engineers, or (c) developed by a corporate PR-department. This requires a certain amount of transparency by those who hold the concentration of expertise - industry. Understanding the fundamentals of *what* we are looking to regulate or not regulate is critical, as is understanding whether the desired outcome is technically feasible. The narrow AI developed today is in many ways the foundation of AGI in the future. Those building and deploying that technology today, need to document the process and the limitations, and be honest about the same, so we better understand them for tomorrow and can build appropriate guardrails around the same.

Second, dialogue on an international level that allows all stakeholders to be more forthright in their positions, concerns, and competing pressures is necessary. This may involve multi-stakeholder collaborations at the outset employing tools such as Chatham House Rules, giving participants some reassurance that their collaboration will not have negative individual consequences. Development of voluntary ethical guidelines for AI, such as those recently adopted by UNESCO, are excellent resources to use to guide the discussion. And international organizations with the expertise to bring together many stakeholders across the world to negotiate critical agreements on standards and rules are perfect for driving the initial collaboration (UNESCO, WIPO, WTO, ISO, IEEE, and others).

Third, where we take that dialogue – or prioritization – comes next. Part of the collaborative process should be agreement on what aspects of regulatory oversight need to be (a) international, (b) national, (c) standards-driven, and (d) self-regulated. We do not need and will not want regulations for all aspects of AI development and use. Over-regulation, just like lack of regulation, will result in poor development, but also difficulty in managing compliance by both companies and regulators (not to mention limiting experimentation and innovation). Similarly, many regulations that currently exist may well apply to the problems we are concerned about with AI. Identifying what truly needs a novel approach will help narrow down the task at hand. To that end, prioritization of issues arising in the development and end-use of AI will help guide this process. Of critical importance at the outset are issues around: Safety and Human Good; Government/Military Use of AI; Human Rights; Competition and Markets; Intellectual Property; Data Rights and Privacy; Education and Labor; and Fairness and Bias.

Finally, we need to address how governments, standards bodies, and companies will address compliance. What mechanisms and bodies need to be in place to ensure regulations and industry standards are met? Who has oversight and what are the remedies? What requires strict compliance and penalties, and what rights do individuals have to seek remedies for harm? How do we encourage ongoing transparency?

Conclusion

Whether or not we need to regulate AI is no longer a question. The biggest issue is what requires regulation, what industry standards are needed, and what should be left to self-regulation based on those priorities that are most important to (a) protect our larger society from negative unintended consequences and misuse, including incursions on the rights of the individual, (b) prevent a race to dominance at the expense of safety or the principal of use of AI for human good, (c) encourage a

competitive marketplace, and (d) allow for open experimentation and innovation. To accomplish this requires all stakeholders be willing to collaborate in an open and transparent manner, to listen to different approaches to the same problem, and to develop solutions that have the flexibility to move with the technology and society as it matures.

Collaboration is difficult and will require years to get right. Many of these processes have already begun and it is critical that every stakeholder come to the table willing to participate in an open and transparent way, particularly those that are primarily responsible for introducing AI and its positive impacts to the world.

WHAT DOES IT MEAN TO BE A DIGITAL NOMAD?

Digital world has changed the conceptualization of so many different terms, and nomadism is not an exclusion. Personal information and identity, arguably two of the most vital pillars of our age, have been shaping the modern understanding of digital nomads. The digitally nomadic are on the go, which explains why digital nomadism isn't characterized by who they are or what they do. It's more of a state, a tension that runs through the lives of individuals who are free to travel because they work in a digital capacity that also refuses to let them go. Understanding such a tension is crucial to completely comprehend the mindset of the digital nomads.

James Brusseau*

TPQ
Winter 2021/22

* James Brusseau (PhD, Philosophy) is author of books, articles, and media in the history of philosophy and ethics. He has taught in Europe, Mexico, and currently at Pace University near his home in New York City.

Nomads experience this radical freedom: they travel because they do not know what they will find. They are distinct from tourists wanting a picture of the Eiffel tower, and different from religious and spiritual travelers seeking redemption at Mecca, or awareness on the Peyote desert of northern Mexico. For all of them, the reason to go was established before they left.

By discovering their trip's meaning after arriving, nomads reverse the conventional relationship between journey and destination. Instead of going somewhere to be there, the destination is only there to justify and orient the going. Practically, there are strategies for achieving this commitment to what does not yet exist. Buying the cheapest air ticket currently available to a capital city on a separate continent is a kind of random location generator, and so a way of going without expectations. Desktop globes can be spun as geographic roulette. Regardless, it is not about the randomness, but that every method initiating nomadic traveling traces back to a single incentive: finding a way to shift the purpose from the beginning to the end.

The Paradigmatic Nomad

A century after her death, Isabelle Eberhardt's story is widely shared as a precursor to today's swirling gender ambiguities. Before she became a symbol though, she was history's paradigmatic nomad because she incarnated what is at stake when the reason for going is filled by the going itself. She was born into minor aristocracy on her mother's side, and into the learned idiosyncrasy of her father, who was a tutor, anarchist, and daring adulterer. Eberhardt studied languages and cultures intensely, traveled to Algeria at twenty, and died there in a flashflood at 27. A life that initiated with privilege and opportunity ended in anguish: she perished nearly toothless and married to an Algerian foot soldier with no prospects.

She "went native," that would be one way to capture her time exploring the geography, religions, and cultures of north Africa. Less sentimentally, she had no return to her past. The French administration in Africa refused to acknowledge her, at least publicly. She had no relatives or home to welcome her back to Switzerland. She had no money to get back even if she tried.

The significance is not that she ended up impoverished and nearly alone, but how she got that way. This is the critical aspect: for her, traveling consumed the traveler. As European women were restricted legally and culturally from journeying alone in northern Africa, she shaved her hair and used clothes to convert her bony frame into that of a young man. From there, free voyaging can be traced through her short autobiographical narratives recounting the trip's cultural and religious explorations.

An extensive list of transformations accumulated. Besides briefly becoming a man in appearance and in name, Si Mahmoud Saadi, she also exiled herself from expatriate gathering places. She adopted the local dietary habits. She learned the rules of desert survival. She experimented with a series of Islamic faiths and ultimately adhered to a remote and mystical sect of Sufism where her devotion and prominence as a believer was sufficient to warrant an assassination attempt by a rival sect.

"Today's nomads arriving with computers and credit cards instead of journals and cash are different at that critical moment. For them, the question of economic survival is more about calculation than adaptation and hope."

Studies produced by Stanford University have revealed the potential for online avatars to modify the real-world behaviors of their users. The idea is that users imitate and then become their avatars instead of the other way around. What is most notable about Eberhart's travels is that she was her own avatar. To explore Africa, she needed to project herself with modified dress, habits, and beliefs. It may be that initially those superficial presentations disguised who she was or stood in for her like a character she played, but eventually the role absorbed her.

About personal identity, Nietzsche prominently declared that under every mask, there is another mask, meaning that every psychological attempt to divine our true selves reveals a superficial representation hiding something else underneath. Nomadism accepts the logic but runs it the other way: on top of every mask, another can be placed, and that presented self becomes the genuine self. It becomes an identity as true and authentic as any other that can be found underneath, or before, or left behind.

Ultimately, what makes Eberhard paradigmatic is that she exposed herself to her voyage. She allowed the habits and values that the trip required to literally define who and what she was. The journey – not the person taking the trip – was the dedication. As it happened, the trip went poorly, and while better outcomes are always possible, her experience also teaches that there are no guarantees, there is no certainty that the emergent identity will be broader or wiser or happier. What comes next may well be narrow and miserable. There is no way to know until it is too late.

The defining characteristics of the nomad:

- The reason for traveling is not knowing what to expect
- The trip's purpose is determined after going
- The traveler's personal identity absorbs the trip's demands

Obstacles to Nomadism Today

Before digital reality streamed employees around the globe, traveling to a different place for an extended stay required getting a job there. The project sunk travelers into the fouler depths of their new city: the exploitation of off-the-books waiters paid in cash at the end of the night, the informal merchandising, the illicit services. There was usually personal debasement, but it always had the virtue of providing valuable lessons about the locals, about who they were, how they acted, what they wanted. There are few more effective ways to gain exposure to others than to be humiliated by them.

Getting the work visa for legitimate employment proved nearly as enlightening. A few passport-sized photos and a gritty lawyer who knew how to expedite the paperwork got the process started. The requisite bribery in some cities cost more money and patience than in others, but once the oily logic of the particular bureaucracy was learned, the experience always proved worth the effort not only by providing a step out of exploitation, but also because the habits of the bureaucrats typically illuminate a culture's unwritten rules.

Today's nomads arriving with computers and credit cards instead of journals and cash are different at that critical moment. For them, the question of economic survival is more about calculation than adaptation and hope. Part of the reason digital nomads congregate in Porto and in Lisbon is that Portugal combines the decadence and cost-of-living of a third world-country, but with low crime, and a resident visa exempting all income earned outside the country from taxation. There is an upfront cost to be paid, but it can be a good purchase for those drop-boxing deliverables to the US from a stone-walled, 18th century apartment. With a few additional administrative acrobatics, the result can be nearly zero taxation, which works for most people.

It is also true, though, that foreignness is stripped away from the experience when it is no longer necessary to wring cash out of the new culture. While it is true that nomads do not travel to make money, it is also true that making money is among the most direct routes to nomadic traveling.

A wider divergence between the pre- and post-digital nomads opens on the level of personal information. It used to be that traveling automatically severed identifying data: in the new city there was no one to know about a traveler's previous tastes and experiences. Today, the website PimEyes will take any pictured face and return the URLs where it has appeared. The links can go back years, chaining travelers to homes, jobs, beliefs, friends and lovers. Anyone who goes abroad to experiment with new ways of living, that means, is always only one iPhone snap away from being unmasked, from being returned to the person they used to be. Then there is LinkedIn constantly reminding users of their accrued professional interests. Facebook occasionally sends old pictures into the current feed. So, the attempt to cut away from our own personal information and bend into the rules and practices of a new culture is constantly sabotaged by the threat of who we have been.

> *"The existence of digital nomads, labors against this sentiment, against the idea that we have only one true self, against the idea that we have an ethical responsibility to be true to that oneself, against the philosophy of the twentieth century, and against the social media technology of the early 21st."*

Probably, it is easier than ever to travel, but current technologies also make it proportionately more difficult to depart from our own pasts. So, obstacles to nomadism today are:

- Calculation replaces adaptation at the departure
- Personal information traps travelers in places – and as people – they have already been

Personal Information and Identity Today

In 2017, Gizmodo reported on a multiple personality Facebook user. She existed nocturnally as an escort arranging trysts online through a discrete account under an assumed name. At the same time and on the same platform she maintained a more ordinary, daytime identity associated with her birthname and a long history of posts detailing the kind of user social media companies seek: steady, conventional, predictable.

The split realities kept their distance until her day self and her night clients began

appearing in each other's "People You May Know" recommendations. Besides knotting the dual-track lifestyle and triggering nervous Google searches, the crossed wires left behind this lesson. While it may be true that the split identities eventually collapsed together online, it is also certain that the same technology tracking users' data and binding them to their own pasts can be twisted against itself, at least temporarily. Despite the Facebook demand for authenticity, creative users find ways on the platform to become someone they were not supposed to be according to their own prior personal information.

Distinct platforms will be vulnerable to different strategies. Dating applications can be perverted to suggest partners who would never cross our paths in material life. LinkedIn algorithms can be corrupted by adding marginal biographical information. It is always possible to cruise through social media sites trying to provoke offbeat connections and links. With some serendipity, a freelancing programmer may end up on a project that leads to a corporate post in the Danish welfare state. A feminist paralegal could end up purchasing a headscarf and a one-way ticket to an English-teaching job in Saudi Arabia. Regardless, life-jarring opportunities are out there, and the personal information that all of us wield online can be hacked to tilt the algorithms and produce deviant possibilities.

For humanism, the twentieth century was about authenticity. The time belonged to Martin Heidegger and to the phenomenologists and existentialists who emerged from the horrors of two world wars. They proposed that the human condition demanded personal integrity in the literal sense of indivisibility. The project of philosophy was fidelity to ourselves as individuals whose resilient uniqueness endured through time, and was guaranteed by the fact of dying alone: only I can die for myself, and I must always be true to that one who will die. Though he is unaware of it, these thinkers are the reason Facebook's Mark Zuckerberg converted digital technology into ethics: "The days of you having a different image for your work friends or co-workers and for the other people you know are ending. Having two identities for yourself is an example of a lack of integrity."

The existence of digital nomads, labors against this sentiment, against the idea that we have only one true self, against the idea that we have an ethical responsibility to be true to that oneself, against the philosophy of the twentieth century, and against the social media technology of the early 21st.

Because nomadism today means working against technology from within it, the conflicted reality of traveling life repeats on the level of personal identity. It has never been easier – or harder – to get out of who we are, to disrupt our existences

from the bottom up by connecting with and nurturing unfamiliar tastes, urges, and potentials. It has never been easier – or harder – to go somewhere different and become someone else.

What Does it Mean to be a Digital Nomad?

Pure nomadism is impossible for digitally mobile workers because it requires going native. As a practical matter, the customs, values, and beliefs discovered in a foreign place cannot consume travelers who are maintaining a regular job back home with set hours and expected behaviors.

But the digitally nomadic are on the road, and that duality – at once at home and abroad – explains why digital nomadism is not defined by what someone is, or even by what they do. Instead, it is a condition, a tension stretching across the experiences of those who are freed to travel because they are employed in a digital way that simultaneously refuses to allow their departure.

1. Geographically, the nomadic tension is between calculation and improvisation. No matter where they go, digital nomads need to plan. To begin, time zones need to be calculated since working meetings are going to be required, and if the group sessions correspond with the New York City working day, then one-way tickets to Portugal and Spain will overlap well enough. Istanbul is a stretch. Kuala Lumpur is red-zoned.

   Obviously worldwide clocks can be managed, but keeping the hours straight is not the problem, the problem is that traveling is being managed. Consequently, what makes someone a digital nomad on the geographic level today is life torn between two kinds of departures. One requires accounting, while the other exists precisely because there is no way to know beforehand exactly what needs to be accounted for, or even what planning will be required.

2. Technologically, the lived tension of digital nomadism involves personal information. On one side, every time the phone lights up or an email sends, ones and zeros flow to platforms and data brokers that crunch the numbers and tighten predictions about where users will be, and what they will be doing, and why. For the digital nomad, hardly a shred of working life escapes this force of information-confinement.

   On the other side, today's digital nomads are uniquely positioned to pervert the machines because global digital platforms provide instantaneous

exposure to local experiences. It took Isabelle Eberhardt years to locate the cultural mysteries of north Africa, now even the most esoteric practices find their way onto the universal visibility of the web. Then the strategic use of invented personal information allows the creation of alternative identities. Just as Eberhardt cut her hair and switched the gender of her clothing to commune with foreigners, so too today's digital nomads may create online disguises tailored to gain entrance to even the most unfamiliar and obscure communities.

3. Existentially, the definitive element of the digital nomad is a tension between two ways of conceiving identity. According to the first, I was born a certain kind of person, and therefore I end up going on the road with a computer and working from hostels and rooming houses. Because of how I am defined, I will stay for a time and then move on to another place that could have been predicted to attract me.

   On the other side there is the one escaping any prediction because there is no way to know beforehand who it is that will emerge from the trip. This is the person who lives the customs and values of a foreign place, and by living them absorbs them, and ultimately incarnates them.

The digital nomad is precisely the one who always answers yes to both sides of this question: Does the traveler do the traveling, or does the traveling do the traveler?

TESTIMONIALS

"The neighborhood of Turkey is where important parts of our new future are now being shaped! And, TPQ brings the perspectives that are essential for shaping the right policies for this new future."

– Carl Bildt, Stockholm

"An essential source for anyone seeking informed and objective analyses of key developments in Turkey today."

– Anthony Giddens, London.

"TPQ's compelling, well-written and edited content has always made it good reading. But now Turkey's dramatically changed place in the world makes it must reading."

– Fred Kempe, Washington, DC.

"TPQ is the journal I turn to when I want to understand the world inwhich Turkey lives and operates, a world which I expect my students and colleagues to understand as well. TPQ restored my faith in opinion/analysis journals by its exquisite ability to frame questions, define themes and select contributors. My association with the TPQ process has become a source of pride."

– Gerard Libaridian, Ann Arbor, MI

"Turkish Policy Quarterly (TPQ) is doubtless an excellent source of information and analysis as well as of the visions emanating from the countries of the region. TPQ's highly professional material and attractiveness make it, I would say, a unique publication. I particularly value its sympathetic feature that among its authors you find not only analysts but politicians as well. I cannot imagine studying the regions' politics without reading TPQ."

– Alexander Rondeli, Tbilisi.

"For those interested in Turkey, including Turkish scholars such as myself, Turkish Policy Quarterly has become an indispensable source as a means of getting informed analysis on current affairs. By bringing together Turkish and foreign experts, TPQ offers a highly varied and uniquely rich perspective in understanding Turkey's politics and international relations."

– İlter Turan, İstanbul

NOTES

WHY CHOOSE BETWEEN BUSINESS AND QUALITY OF LIFE?

The MONACO ECONOMIC BOARD seeks and accompanies foreign businesses wishing to establish in the Principality. An important aspect of its role is to work closely with the Monaco Gouvernment, and in particular the Welcome Office, to assist with the setting up process. INVEST MONACO is on hand to help companies every step of the way to ensure setting up a busness in the Principality is a simple and straightforward as possible.

MONACO ECONOMIC BOARD

contact: info@meb.mc

2,ruedela Lüjerneta - 98000 Monaco - www.meb.mc - Tel.:+3779798 68 68

"Over the last 15 years, TPQ has earned recognition and respect in the publishing world. It has managed to create its own intellectual network and accumulate extensive experience."

Ekmeleddin Ihsanoglu, 2017

"Over the years, TPQ has been one of the best sources for debate and analysis on contemporary Turkey, its region, and the country's international partnerships – and rightly so."

Ian Lesser, 2017

"The steps taken to keep pace with the digital world has assured TPQ of continued success in providing informed analyses on Turkey."

Faruk Logoglu, 2017

"In today's complex and turbulent world, TPQ's capacity to offer a vast array of perspectives on the geopolitical issues of the time is a genuine asset for its readership."

Marc Pierini, 2017

"TPQ can confidently lay claim to so much influence on the shaping of Turkish foreign policy in the last one and a half decade through its incisive strategical analysis on a wide variety of topics."

Namik Tan, 2017

"I find TPQ an enriching addition to the contemporary literature, an open forum where the variety of opinion is refreshing, and a source that I can turn to in preparing for talks I am often asked to give."

Ilter Turan, 2017

www.turkishpolicy.com

CERTIFICATE NO: 21626

ISSN 1303-5754

9 771303 575007 11

twitter.com/turkishpolicy facebook.com/turkishpolicy