Briefing note

Digital Transformation and the Future of Democracy

Fundamental Rights in the Digital Era

Data Economies and the Future of the Social Contract

Trust and Public Debate in the Disinformation Age

>> Background

The World Leadership Alliance-Club de Madrid (WLA-CdM) is organising its 2019 Annual Policy Dialogue in partnership with the IE School of Global and Public Affairs, bringing key stakeholders to Madrid for a timely discussion on the implications of digital transformation and Artificial Intelligence (AI) for democracy. The event aims to take stock of the current debate and propose key policy recommendations to ensure that digital technologies and, more specifically, AI systems, are designed, developed and deployed to benefit individuals and societies while upholding democratic values and institutions. The dialogue will offer a unique platform for multi-stakeholder conversations involving representatives of governments, academic institutions and think tanks, tech companies, and civil society, as well as 30+ democratic former Heads of State and Government who are Members of WLA-CdM.
**Objective**

The Policy Dialogue is organised with a twofold objective. Firstly, to offer a multi-stakeholder platform that can stimulate new thinking in response to arising opportunities and challenges of AI and other digital technologies that are relevant for the future of democracy. Secondly, to enable the identification of the most effective policy response, including the establishment of oversight structures, to support AI as a driver for stronger democracies while countering its potential risks.

**Background**

Digital transformation brings along great opportunities for democracy but also enormous governance challenges. Almost no element of our social fabric is spared from the impact of these disruptive technologies, which are rapidly reshaping the way citizens work, live and communicate.

The extent to which governments and corporations succeed in amassing and using relevant data – the means of production of AI – is set to alter the global economy and the balance of power between states, markets and civil society. The rise of AI is reshaping the geopolitical and societal orders in ways researchers are only beginning to examine.

The use of AI may pose a challenge to democracy, but, if handled correctly, it can also bring more and better democracy. Democratic governments simply cannot afford to lag behind; they must govern the digital game before it governs us all. The digital is political and therefore requires a political response: How can we anticipate the fast-changing world of AI and reap the benefits while countering the risks it poses to democracies?

Not only are digital technologies reshaping global politics, but they are affecting the very core of domestic governance too. The use of social media, bots and automated systems to interfere in electoral processes is but an example of this. The combination of the digitalization of public debate together with highly sophisticated means of election hacking can unsettle the legitimacy of democratic institutions and with it the very foundations of the liberal order.

Engaging with actors across the globe, including the tech industry, will be of the essence to garner consensus around a new world order shaped by exponential digital transformation, but liberal democracies must first protect themselves by facing up to the challenge of redefining an increasingly contested system of democratic governance in the age of AI.

The Policy Dialogue on AI and democracy seeks to move the conversation on AI beyond tech and into the democratic governance arena. The assumption is that the digital is political as coined by Jamie Susskind in his book “Future Politics – Living Together in a World Transformed by Tech” (2018). By bringing together experienced politicians, tech companies, academic researchers, and civil society representatives, the organizers are aiming to promote ‘multi-stakeholderism’ in the articulation of informed policy proposals and solutions that can effectively turn the design, development and deployment of AI into a driver for democratic innovation and renewal at a time when widespread dissatisfaction with the present system and uncertainty about the future are seriously affecting public trust.
Expected Outputs

- **Multi-stakeholder engagement:** A common understanding on the potential risks and benefits of AI is reached.
- **Call for action:** Common positions and key policy recommendations on digital technology/AI and democracy are identified in view of raising awareness and informing the actions of multiple stakeholders on the matter.
- **Action points:** Two-three initiatives for concrete action are taken forward as projects by WLA-CdM and partners.

Rationale

The overarching questions guiding the policy dialogue are:

- To what extent do digital technology and AI strengthen or threaten democracy?
- What kinds of policy responses are required to address the increased use of AI and its multifaceted implications for democracy?
- How can political leaders advance such multi-dimensional policy responses?

The following sub-themes will guide the three action labs in which concrete policy recommendations will be discussed and agreed upon:

A. Fundamental Rights in the Digital Era

AI can be used to strengthen democratic governance and institutions as long as the design, development and deployment of intelligent systems is done in a manner that upholds fundamental rights and core democratic values. Additionally, it can empower societies by enabling the creation of programmes that bring progress to humanity. However, the exponential autonomous collection, processing, management and distribution of data –the means of production of AI– may be significantly invasive and pose a major challenge to traditional definitions of privacy, further deriving in an infringement of the fundamental rights of freedom and equality.

Massive data collection has increased the possibility of repressive surveillance on an unprecedented scale. Pervasive tracking can give platforms important information on user behaviour, which can end up determining the interest they pay on a loan
or their access to a job listing. Data brokers can place individuals in high-risk classifications based on their search history, further enabling discrimination. AI systems are also overhauling key sectors such as the insurance industry, which can lead to personalized pricing based on indicators that are little but proxies for factors that would otherwise be illegal to consider, such as race, sex, poverty or genetics. Furthermore, decisions that have traditionally been made by governments because of their nature and impact on human lives, currently belong with tech elites that do not play under the same rules. This unprecedented concentration of power in corporate hands raises urgent questions pertaining to the legality and legitimacy of their actions. The privacy challenges that arise from the Big Data society require a deep debate over the ownership and treatment of information. Special attention must be given to the issue of transparency and accountability surrounding the collection and use of data by private actors.

Increasingly too, decision-makers are turning to AI to render governance more effective and efficient and improve their public policy responses. Good practices surrounding AI-powered public service delivery abound. However, the adoption of automated decision-making systems by governments raises important challenges related to transparency, reliability and accountability. The rise of black box algorithms can perpetuate bias while hindering political responsibility.

*Should individuals exert more or total ownership of their own data, or can a balance be struck with corporations and governments that would allow for the collection and use of data for targeted service delivery while protecting the fundamental rights of the individual? How can governments ensure that online service providers refrain from using or manipulating data for their benefit or that of third parties? Whose responsibility is it to guarantee that digital technologies are not discriminatory? What regulatory measures should governments adopt to ensure citizens’ fundamental rights are guaranteed? How can automated decision-making be regulated to guarantee accountability?*

**B. Data Economies and the Future of the Social Contract**

Promises linked to the development of the data economies are only matched by the already existing risks: rising inequalities; power concentration; and undermining of the democratic systems. Current asymmetries of power between the few tech corporations and democratically-legitimized national governments pose a significant challenge to our political system, at times rendering existing governance structures and institutions ineffective or obsolete. Ever-increasing data flows fuel economic growth, yet the distribution of its benefits poses significant questions. Wide adoption of AI across public and private sector allows for efficiency gains, at the same time exacerbating current lines of socioeconomic and political divides, shaking the fundamentals of the post-WW2 liberal social contract. A new level playing field is needed to secure inclusive, beneficial and democratic growth. The purpose of this group is to reflect on the policy responses that can facilitate new institutional arrangements on the national level.
C. Trust and Public Debate in the Disinformation Age

Digital technologies have opened new channels of communication and coalition building that allow for direct interaction between political leaders and citizens. They have also created a space for the expression of political ideas that might otherwise not find their way to the political debate. Digitalization has lowered the barriers for citizens to engage in nation-wide political conversations. In countries where traditional media cannot veer free from government restraint, the use of internet and social media provide an alternative outlet free speech.

However, digital technologies also bring about new information challenges such as the extensive use of persuasion architecture. Online anonymousness, zero-cost publishing and content retransmission are favouring the propagation of political messages, hate speech, extremist and polarizing ideas that would meet more hurdles and receive less attention in the non-digital world. Algorithms, whereby search engines and newsfeeds prioritize content based on each user’s profile, have created echo chambers that push online citizens away from multi-faceted analysis into ideological one-sidedness, blocking the construction of public debate. AI-driven technologies take advantage of our identities using algorithms to create specific targeted content to perpetuate existing bias; this is further reinforced when such content is embedded in deep fake materials or fake news. Controversies surrounding the alleged malevolent viral circulation of fake news during the 2016 presidential election in the US and the 2016 Brexit referendum are just two recent examples of how democratic systems can be affected by attacks on information integrity in the digital environment.

How can democratic governments use AI to combat fake news in order to support information integrity? How can AI be used to identify and stop deep-fake videos before they spread and what role should governments and digital information companies play in this regard? How can AI be used to create a transparent, reliable and productive interface between the government and its citizens? Should governments regulate the internet in order to promote information integrity or does regulation inevitably
Working Group A

Fundamental Rights In the Digital Age

Introduction

This Background Paper was prepared collectively by members of the Working Group on “Fundamental Rights in the Digital Era” steered by Nicolas Mialhe (The Future Society) ahead of the Policy Dialogue. It is intended as a starting point for the Action Lab discussions that will take place during the Dialogue and as a basis for the final “Call to Action”.

Working Group Members were Helen Darbishire (Access Info Europe), Marc-Antoine Dilhac (University of Montreal), Shawnna Hoffman (IBM), Moussa Kondo (Now Generation Forum), Hans Kundnani (Chatham House), Luz Amparo Medina (OEI), Ralph Müller-Eiselt (Bertelsmann Stiftung) Prime Minister Iveta Radičová (WLA-CdM), Prime Minister Hanna Suchocka (WLA-CdM) and Cedric Wachholz (UNESCO). The ideas expressed in this Background Paper do not necessarily represent the views of individual Working Group members.
Introduction

From accelerated innovation and sharing of ideas to improved products and services across industries, the Digital Era has created clear benefits. However, it also brings a unique set of risks that policymakers worldwide must urgently understand.

A powerful dialogue around the economic and socio-ethical implications of the Digital Era has already begun to spring up throughout the world. However, now more than ever, it is time to expand this ongoing dialogue and approach it through the perspective of fundamental rights. There are many multi-faceted questions that policymakers worldwide must now confront, including:

- What do fundamental rights in today’s digital era look like and are they under threat?
- What should they look like?
- Are they still the indispensable foundation and guard rail as such, or do they need to be augmented?
- Do the values and notions embedded within them fit for the transformations unfolding?

This background paper was developed by the ‘Fundamental Rights in the Digital Era’ working group at the World Leadership Alliance - Club de Madrid 2019 Policy Dialogue. It aims to explore the complicated questions above by triggering a much needed and time-sensitive global multi-stakeholder dialogue that will create a path forward to place fundamental rights at the heart of today’s most pressing challenge: to reap the opportunities of the Digital Era while also mitigating its risks.

Firstly, the paper frames the conversation by setting the parameters - defining what exactly is meant when we say ‘Digital Era’ and which fundamental rights will be covered in this scope. Secondly, the opportunities will be unpacked – exploring how the Digital Era could positively impact fundamental rights, helping protect and enforce them across societies worldwide. Thirdly, the risks will also be enumerated – looking at how the Digital Era could violate fundamental rights and the risks ahead. Finally, the background paper will provide some practical recommendations for the way forward, highlighting the need for a global and inclusive dialogue to create an updated set of fundamental rights – a set of rights fit for the Digital Era.

Context

What is the ‘Digital Era’?

The Digital Era is characterized by significant advances in new technologies (i.e. Nanotechnology, Biotechnology, Information and Communication Technologies, AI, Blockchain, and more) as well as social and economic transformations worldwide. It has three key dynamics:

1. the growing importance of data in shaping products, services, and business models;
2. greater returns to scale; and
3. network effects that give rise to the platform economy.

These three dynamics converge within multi-sided business models that have become prevalent with the rise of the digital economy. Pervasive connectivity, open innovation, automation and data-driven decision making are just some of the common trends.

In reality, nearly every aspect of our everyday lives has been affected or transformed by the Digital Era to some extent. Work, education, family, health, shopping, leisure, and communications are just a part of the many aspects that have been revolutionized in the past few decades. Moving forward, the revolution will truly touch all dimensions.

This paper sheds particular attention on the impact of the increasing progress and deployment of automated decision making (ADM) systems and AI technologies, as these specific developments have become some of the most powerful narratives of our period. These technologies have their own set of opportunities...
and risks. For instance, the use of ADM systems and/or AI technologies has the capacity to unlock enormous potential in societal, political, economic and cultural processes - including better personalization of products and services, easier access to public goods, fairness at scale, individual empowerment, and rapid progress towards the United Nations Sustainable Development Goals. They also could raise new risks like increasing inequalities, technological unemployment, algorithmic bias, manipulation, loss of agency, threats to privacy and security, and many more. Furthermore, as the world economy transforms, lagging behind in adopting AI and emerging technologies can mean a widening economic and human development gap between countries, people, and companies.

Which Fundamental Rights?

In light of the socio-economic changes and technological breakthroughs of the Digital Era, it is essential to place fundamental rights at the center stage. While there are many different conceptions of fundamental rights, this background paper builds from two key sources: the Universal Declaration of Human Rights (UDHR) and various publications around new technology ethics principles.
The UDHR declaration is underpinned by four key universal values: human dignity, freedom, equality and solidarity. Most of the publications on technology ethics also take into account these same universal values in the context of a world in which AI adoption is exponentially growing.

The Berkman Klein Center for Internet and Society at Harvard’s visual mapping of ethical and rights-based approaches to principled AI pinpoints eight shared themes across key publications: accountability, fairness and non-discrimination, human control of technology, privacy, professional responsibility, promotion of human values, safety and security, and transparency and explainability.

These principles can be identified in key initiatives on ADM systems and AI guidelines and principles. For example, the OECD AI Principles, developed by a group of multi-stakeholder experts (including The Future Society), include values-based principles which serve as recommendations for the “responsible stewardship of trustworthy AI.” The G20 has drawn on the OECD’s principles to include provisions for “Human-Centered Artificial Intelligence” in its June 2019 G20 Ministerial Statement on Trade and Digital Economy. The European Union’s High-Level Expert Group on AI has put forward recommendations for safe and ethical AI development in Europe. The Montreal Declaration for a Responsible Development of Artificial Intelligence has offered a rights-based approach to ethical guidelines for the development of AI. Other civil society initiatives such as Bertelsmann Stiftung’s Algo.Rules provide the basis for ethical considerations as well as the implementation and enforcement of legal frameworks.

Opportunities

Human Life and Dignity

A principal fundamental right is that which protects a human being’s right to life, securing a human’s dignity and respect for his or her physical and mental integrity. The Digital Era ushers new opportunities to close gaps in life quality.

Technological breakthroughs in the Digital Era have the power to save human life; in healthcare, for example, new technologies like AI can increase access to quality healthcare by improving the productivity of doctors, accuracy of diagnostics, and efficiency in services. Furthermore, AI and big data analytics enable predictive capabilities and resource efficiency to expand access and inclusion of clinician services into remote areas at a lower cost. Personalized and precision medicine, more accurate and faster diagnostics, and accessible health apps increase access to quality medical care for millions. Chatbots offer 24/7 free therapy, accessories monitor biometric data in real time, and robotic devices improve surgical outcomes.

The socio-technological trends of the Digital Era also have the potential
to dramatically improve the quality of human lives worldwide. Digital technologies enable increases in innovation through greater market competition, lowering barriers to market entry for smaller actors, and increasing consumer welfare through innovative products and services. This can empower more small businesses, connect people, and help support fulfilling work and sustainable living.

Education

Education is a second fundamental right and includes access to vocational training and lifelong learning. The Digital Era brings new opportunities to expand access to high quality education including to underserved populations across the world. There is a major role for AI systems specifically to promote personalized learning methods and curriculums, which create content based on students’ preferences and performance, along with digital learning applications and tools to make teachers’ work more efficient. For example, at the 2019 Global Governance of AI Roundtable (GGAR), multi-stakeholder AI experts discussed how to apply AI to solve gaps in education worldwide. This approach can revolutionize our current ‘one-size-fits-all’ education systems and help ensure individuals across the world have access to high quality education, fit for their own unique personalities.

Risks

Privacy

Although the opportunities are huge, the Digital Era can also clash with fundamental rights. One of the rights most negatively affected by the Digital Era is that linked to notions of privacy. AI applications, for example, can be used to track individuals across devices. Although the consequences of such tracking can result in improved products and services for consumers, it can also impact people’s privacy, especially when an individual’s data is collected or processed without their consent or awareness.

In a commercial setting, this may create a strong information asymmetry that results in skewed control dynamics between data collectors and individuals providing their data. In a public sector setting, this can bring rise to ethical questions related to surveillance capitalism.

Policymakers, in today’s Digital Era, must take these socio-technological transformations in mind and build (or rebuild) data rights which protect peoples’ privacy.

Freedom

With the introduction of the Internet, the right to freedom (i.e. freedom of thought, conscience and religion; freedom of expression and information; freedom of assembly and of association; freedom of the arts and sciences) is living a never seen expansion. In many ways, the Internet and other technologies have democratized it, breaking economic, geographical and age barriers. The Arab spring was just an example.

On the other hand, however, the Digital Era severely challenges the right to freedom. Information manipulation, micro-targeting, risk profiling, and so called “aid to decision” tools require skills and understanding by the public to serve as conduit to emancipation and trust, as opposed to alienation and diffidence. Appropriate governance frameworks and protocols must be designed and implemented to empower individuals and communities freely make decisions and participate as citizens. Achieving that goal is rendered more difficult by the cultural and political divergences between regional and national value systems in addressing tensions between free speech, respect for others, and individual rights. A global consensus is emerging around the fact that automated decision-making processes affecting a person’s life, quality of life, or reputation must be
transparent, fair, accountable, and accessible to individuals.

More than just addressing privacy concerns, policymakers must provide a secure space for individual freedom and agency while participating (living, working, and learning) in the Digital Era.

**Equality and Non-Discrimination**

Lastly, the Digital Era threatens to contribute to widening income inequality, both within and between countries. Historically, the introduction of new technologies - fire, wheels, pulleys, the printing press, running water, the steam engine, electricity, the telegraph, phones, and computers - have separated those populations with access to those without it. This time is no different.

Studying deeply the case of AI, the presence of algorithmic bias in AI applications may reproduce and further aggravate social marginalization for underrepresented groups. As AI applications scale across key sectors, the impact of discrimination has the potential to grow. Further, the concentration of wealth distribution within AI industrial value chains combined with the accelerating pace of job automation may dislocate middle classes, generating systemic instability. Meanwhile, lack of explainability and basic transparency can make it difficult to achieve the needed social trust in order to take advantage of its benefits. Countries face the pressing challenge of addressing the risk of AI systems creating and reinforcing exclusion and, or discrimination based on social, sexual, ethnic, cultural, religious, and tomorrow genetic or cognitive differences.

Recognizing the threat to equality and harmony, policymakers must work towards institutionalizing the need for diversity and inclusion in all stages of technology development - from the production to the implementation - in hopes that the benefits are evenly distributed across peoples.

**The Road Forward**

The road forward means taking into account the opportunities and risks of the Digital Era through the lens of fundamental rights. While we do not necessarily need new digital rights, we should be prepared to revisit and revise the existing fundamental rights catalogue, which was defined prior to the Digital Age. We must first ensure fundamental rights are safeguarded in light of the socio-technical transformations unfolding and, second, adapt them only if need be.

Key in this is realizing that the impact is not distributed equally and the perspectives around how the opportunities and risks should be addressed vary. Some individuals and groups are affected more strongly than others, both negatively and positively. And, at times, certain elements can positively impact the enjoyment of a fundamental right by some while adversely impacting it for others.

Consequently, governance should focus on societal impact, starting from the situation of the most vulnerable, as a basis to then build the right balance between misuse and “missed” use of digital systems. The scope of governance should be on the white box vs. black box spectrum (technology-aware), depending on an AI or other ADM system’s potential socio-technical risk and, especially, the extent to which it can lead to discrimination against individuals or groups of individuals. Systems should be assessed in terms of their social impact based on criteria such as:

a. the number of people affected by the decision-making process and their ability to obtain redress realistically;

b. the degree to which people could be disadvantaged by the decision-making process; or

c. the political and economic power of the system operators. For example, ADM systems used in automated production lanes or similar environments might not require the same scrutiny as ADMs used in the public sector or by credit inquiry agencies.


First of all, we need to enforce and strengthen the existing legal framework: AI and ADM does per se not require the establishment of new fundamental rights. Instead, procedures for enforcing existing individual and collective freedoms and rights must be strengthened in response to the new risks of the digital era.

Steps are already being taken by policymakers on a national and cross-national level to protect fundamental rights in the digital era. For example, the EU Parliament has adapted the EU Charter for Fundamental Rights by adding ‘protecting personal data’ as a fundamental right, stating clearly that citizens should be able to decide freely how to use their own personal data to avoid abuse. On the other hand, many developing countries perceive the innumerable benefits that technology provides to assure access to other rights - such as water, food, or security - more urgent.

Heads of governments around the world are recognizing the need to work towards a common set of global principles to shape the norms and standards that will guide the development of emerging technologies. In a recent speech to the United Nations, UK Prime Minister Boris Johnson highlighted the need to embed rights of freedom, openness, and pluralism in the design of new technologies from the outset in order to safeguard peoples’ rights.

International organizations (e.g. UN, ITU, OECD, GPAI) and supranational government bodies (e.g. EU, African Union, Nordic-Baltic Region, G20, G7) are coordinating policies and pooling resources across countries to devise and implement digital strategies. These provide platforms for coordinating actors to achieve shared objectives and manage global challenges. Policies at international levels can also help to raise or avoid a ‘race to the bottom’ in market incentives, regulation, practices, standards globally.

More needs to be done however to reach a global “regime complex” to align the rise of emerging technologies with fundamental rights. Together, we need to work towards a shared and inclusive vision of fundamental rights in the digital era. Given the diversity of cultures and perspectives of peoples around the world, this is not an easy task. A global civic forum on AI ethics - bringing together citizens, experts, public officials, industry stakeholders, civil organizations and professional associations - could be a way forward.

For example, a joint initiative by UNESCO, The Future Society, the University of Montreal, the Observatory for the Societal Impact of AI, and MILA will host an open global forum for this timely discussion to take place over time - in hopes to collectively achieve an equitable, inclusive, and ecologically sustainable digital world.

Inclusion of all cultures will be critical to mitigate harms and ensure the Digital Era benefits society broadly. With this goal in mind, principles will be interpreted in a coherent manner, while taking into account the specific social, cultural, political and legal contexts of their application.

Approaching the transformations of the Digital Era through the perspective of fundamental rights can serve us all, including political representatives, whether elected or named, whose citizens expect them to take stock of developing social changes, quickly establish a framework allowing a digital transition that serves the greater good, and anticipate the serious risks presented by the Digital Era.
References


2. For a mapping of ‘Principled Artificial Intelligence, see: Berkman Klein: https://ai-hr.cyber.harvard.edu/images/primp-viz.pdf


4. Rules, an initiative by TheBertelsmann Stiftung, is a catalogue of 9 formal criteria developed by experts for enabling the socially beneficial design and oversight of algorithmic systems. For more, see: https://algorules.org/en/home


7. For more, see: https://ethicsofalgorithms.org/2017/06/02/calculating-participation-how-algorithmic-processes-impact-opportunities-to-be-part-of-society/

8. For more, see https://www.youtube.com/watch?v=XaN-MbGV4dY&feature=youtu.be
This Background Paper was prepared collectively by members of the Working Group on “Data Economies and the Future of the Social Contract” steered by Maciej Kuziemski (IE Center for the Governance of Change) ahead of the Policy Dialogue. It is intended as a starting point for the Action Lab discussions that will take place during the Dialogue and as a basis for the final “Call to Action”.

Working Group Members were Chioma Agwuegbo (Now Generation Forum), Edward Corcoran (BBVA), Otto Granados (OEI), Prime Minister Zlatko Lagumdzija (WLA-CdM), María Isabel Mejía (CAF), Gianluca Misuraca (European Commission), Teemu Roos (University of Helsinki), Christoph Steck (Telefónica) and Prime Minister Jigme Thinley (WLA-CdM). The ideas expressed in this Background Paper do not necessarily represent the views of individual Working Group members.
Introduction

The promise of new digital technologies to improve lives is compelling. However, decision makers in government, business and society as a whole need to tackle the risks that this epochal change of digitalization will inevitably bring. We face the important task of seizing new opportunities that technologies present for societies and economies, while mitigating the risks. Filter bubbles, hyperpolarization are linked to populism and rapid changes in the social ordering that disrupt democratic process and contribute to fragmentation of modern society. The cornerstone of these processes are data and algorithms. The key question is how to govern these new technologies to act toward a human future while still respecting and protecting the values of liberal democracies as we know today.

While oldest questions on the foundation of democratic societies based on the idea of human dignity are still the baseline for revisiting those social agencies born in an analogic era, many existing rules, frameworks and processes no longer fit to resolve these challenges. We have to build a new architecture and develop new ideas on what we have already achieved. We cannot start from a blank sheet.

The Working Group has reflected in particular on the following questions:

What government levers and public policies can be used to secure just distribution of the benefits of the digital economy?

What long term educational strategies can nation states adopt to prepare its population for what the future labor market holds?

How to democratically debate and decide about emerging technologies under filter bubbles, hyperpolarization, and populism?

Equality and Non-Discrimination

The group has identified three broad areas that require utmost attention: rethinking the models of growth; securing the competencies of the future in education and employment; and upholding democratic institutions.

A) Models of Growth

Even as global labor productivity grew by 74% between 1973 and 2013, worker compensation grew by only 12.5% over the same period, according to the Economic Policy Institute. While a relatively recent slowdown in global productivity has contributed to the reduction of the real value of salaries, other forces are also at work: rapid technological change, evolving market structures, and globalization, all of which began impacting distribution patterns even before the productivity slowdown. As a result, levels of social inclusion have either deteriorated or remained unchanged in 20 of 29 advanced economies around the world over the past five years, according to the World Economic Forum’s Inclusive Development Index 2018.

During this period, digital business models have flourished, as evidenced by the large number of technology firms, such as Apple, Facebook and Microsoft, who today top the list of most valuable publicly listed entities. For these firms, key features of digital markets such as network effects, economies of scale, and the collection and use of data, have been integral to their growth. However, the concentration that has emerged across many markets – whether in search, mobile operating systems, social networks, or e-commerce - raises questions as to whether competition remains fair and markets are still open to innovation from new players. Furthermore, there are increasing worries as to whether some digital markets, with their dependence on data, are leading to the exploitation of users’ privacy.

At the same time, profit diversion to low-tax jurisdictions provides an advantage for global service providers over local rivals. Tax arbitrage allows global firms to offer their services at lower rates, making the playing field uneven and making it even more difficult for smaller and local companies to compete. Furthermore, this competitive disadvantage of local companies negatively impacts their profitability and thus diminishes their tax contribution to public
finances even further. In the long run, greater profit diversion strips local governments from investment and employment opportunities.

This double tax effect results in an unfair redistribution of taxes in favor of digital producer countries, those where digital leaders are settled, namely US and China to some extent, and to the detriment of digital user countries, such as Europe, Latin America, Africa and the rest of Asia. The unfair geographical distribution of taxes adds more pressure on national budgets at the same time as digitalization demands governments to support the digital transformation of public services. Tax sustainability and social welfare financing are thus becoming an ever increasingly interlinked matter in the wake of digital economy.

B) Future of education and employment

Education is no longer something that you do at a specific institution for a specific period of time to obtain a certification. It is becoming a lifelong learning journey, where practical skills, and the ability to adapt quickly can be more relevant than traditional qualifications. Growing accessibility, affordability and portability of knowledge - including online certification and open access movement - introduces competing means of qualification, especially in rapidly progressing fields such as software development. At the same time, those most affected by the digital transformation (lower skilled workers) tend not to receive online education nor have many training opportunities at the workplace.

How to seize the opportunities and counterbalance the risks brought about by such process and the relationship between different governance stakeholders is a crucial problem to address, and requires a paradigm shift in how the public, private and third sector traditionally operate by unleashing social innovation and new forms of policy making based on collaborative governance mechanisms.

The digital economy is significantly reshaping labor markets, leading to the emergence of new forms of inequality, lower job satisfaction, and lasting unemployment. It is equally a priority to prevent technological change from being accompanied by such phenomena; thus, it is important to rethink what is the substance of meaningful employment in the twenty-first century. Governments, companies, and educational institutions must be able to anticipate future labor market needs, and prepare their constituencies accordingly.

This increasing mismatch between the labor market demand and existing educational offer needs to be addressed through the creation of new curricula and on-the-job training programs. At the same time, workers whose professions are being phased out due to the changing nature of the economy, need to be offered an opportunity to reskill in order to find meaningful employment. Despite the overwhelming agreement and political support for such rhetoric, it is now key to move to concrete actions of retraining, mindful of existing best practices, as well as increased limitations of the state budget in many developing countries.

C) Upholding democratic institutions

The prosperity of liberal democracies is contingent on the ability to align the technical, economic, and political factors that are hollowing out and dismantling existing institutions and procedures. What is at stake are free and fair elections; the active participation of the people, as citizens, in politics and civic life, and the protection of basic human rights of all. One of the most important frontiers is the urgency to challenge the powerful global monopolies that threaten competition and innovation while also becoming curators and moderators of public and private spaces without democratic accountability.

Regardless of a dominant narrative of exceptionalism of big tech companies, governments still possess powerful tools to set the limits of what is and is not desirable, either in the form
of regulation on the national level, or through steering R&D resources towards public benefit technologies via higher education, and funding innovation as well as regional development.

Heightened sophistication of digital technologies, such as microtargeting or machine vision, creates the temptation for governments to anticipate the needs, monitor behaviors and manipulate its citizens. Such situation requires new levels of transparency to hold governments accountable. Yet, a new consensus is emerging that transparency alone is insufficient and without careful contextual considerations, can lead to unanticipated outcomes. In order for transparency initiatives to contribute to greater accountability, information describing how government commissions and uses emerging technologies (i.e. automated decision systems or facial recognition software) needs to be accessible, and a clear auditing and harm remedy scheme need to be put in place. Policy initiatives such as the EU’s High Level Expert Group on AI are a prime example of considering both sides of the equation - the governance of technologies, and the governance with technologies.

**Policy Recommendations**

The urgency of the identified challenges requires immediate and coordinated policy action from all relevant stakeholders - including public and private sector, civil society and academia.

**Inclusive growth**

**Taxation.** It is necessary to speed up OECD’s efforts on the inclusive global solution to taxation in the digital economy, while at the same time empowering national governments that are considering adopting own frameworks. These efforts will be meaningless without more responsible tax behavior by digital service providers, who need to pledge their commitment to the fair local taxation, linked to the service provision in the given territory.

**Competition and consumer protection.** To create the right environment for innovation and trust in the digital economy, both competition and consumer protection policies need to be modified for the digital age. Antitrust authorities need to be more proactive in digital markets and ensure their toolkit is up to date. Regulation can also play a role by returning control to users: there is a need for robust privacy frameworks, as well as regulation to unlock data monopolies. Users should be able to share in real-time their data held in one firm with another. This would put users back in the driving seat while fostering new data-driven innovation. European Union’s General Data Protection Regulation (GDPR) has already become a strong point of reference for policymakers worldwide.

**Social impact bonds.** One powerful tool to mobilize private capital for public good are social impact bonds - a financing tool where creditors fund improved social outcomes for underserved communities - which in turn can result in public sector savings. These bonds are contracts signed by local governments, banks, and foundations that serve a variety of functions, such as helping the homeless find shelters or rehabilitating young criminal offenders. The repayment of related financing depends on how successful a program has been in achieving defined targets, and backers often recycle that repayment back into other projects. Invented fewer than 10 years ago, social impact bonds are used by a wide variety of stakeholders, including the International Red Cross.

**B-corps.** Perverse focus on returns and growth, a systematic problem in the twenty-first century, can be mitigated by incentivizing companies to declare themselves Benefit Corporations (B Corps), for-profit entities that align their corporate interests with those of the environment and the society. Companies operating under these statuses must report to shareholders on how they are balancing conflicting interests and must carefully measure societal impacts of their activities. Benefit corporation legislation has been passed in more than 30
jurisdictions in the US, while Italy has also adopted related provisions; Patagonia and Kickstarter are examples of the thousands of registered benefit corporations.

Skills of the future

Empowering neurodiversity. While the exact nature of how the labor markets will develop is unknown, the workplace of the future will make sure that each person can play to their strengths. Autistic employees may need specific equipment, such as headphones to reduce auditory overstimulation. ADHD people may also require minor adjustments to their work environment in terms of having quiet places to work and flexibility in their work schedules.

Reskilling. To cater to the increasing demand for technical expertise, it is necessary to implement wide transition plans that will allow workers whose professions are being phased out to find meaningful employment. Inclusivity and diversity need to be strongly promoted in order to make sure that the developers of new technologies are able to identify the needs of underrepresented and vulnerable populations.

Digital literacy. General education on all levels needs to raise awareness of the opportunities and hazards related to the use of emerging technologies. This should include media literacy - to be able to assess the trustworthiness of the sources; cybersecurity - to establish privacy enhancing routines; and ethics - to foster healthy conversations.

Democratic procedure

Digital political ads transparency. Granular, data-driven insights about voters have become a standard in political campaigning, disrupting the public fora with micro-targeted messaging and misinformation able to sway the election results. To prevent further interference in the electoral processes, it is key to convince major advertising platforms such as Google or Facebook to adopt binding global standards banning microtargeting and dark posts in political campaigning.

Algorithmic impact assessments. In the wake of an ever increasing adoption of automated decision systems by public agencies, it is crucial to develop global best practices that will adopt a multidisciplinary approach grounded in empirical data to help assess the way these systems are commissioned, built, deployed, as well as their societal impacts, including wider socio-technical context and potential unintended consequences.

Public consultation standards. Rapid socio-economic shifts pose the risk of leaving the most vulnerable communities behind. At the same time, aggressive adoption and diffusion of tools is already causing “techlash,” a growing hostility towards the Silicon Valley model of innovation that epitomizes what Shoshana Zuboff calls “surveillance capitalism.” In the era of the crisis of trust in the public sector, governments that want to reconnect with their constituencies must adopt public consultation standards that collectively reflect about the directions of sociotechnical development and the red lines that should not be crossed.

Practical Examples

Government-led upskilling. A public-private partnership between the University of Helsinki and a Finnish IT company Reaktor has resulted in a development of an online course that provided an entry level training on Artificial Intelligence, with an ambitious aim of drawing in as much as 1% of the entire Finnish population. This goal has now been reached, and the course is being rolled out in Sweden, Germany and Estonia, serving as a testament to the government’s ability to achieve transformative effects through education.

Human rights first AI Strategies. In 2018, the member countries of Digital 9 (D9) developed and agreed on a series of general objectives on the application and use of AI by national governments. In turn, as of 2019, this group of countries formed a working group to share and generate knowledge on the subject, such as frameworks for responsible use, impact analysis on the development of
algorithms and models, followed up by wide public consultations.

**Trust by design.** Brazilian government has established Transparency Portals, which empower citizens to find data about the public sector’s salaries and spending. Seoul’s OPEN system has facilitated citizen complaints about fees being charged illegally by government officials. Armed with information, citizens can monitor gaps in the delivery of goods and services that may indicate corruption.

## Further Reading

A guide to using artificial intelligence in the public sector


Artificial Intelligence and Human Rights

[https://ai-hr.cyber.harvard.edu/](https://ai-hr.cyber.harvard.edu/)

Automation Readiness Index – The Economist – Intelligence Unit

[https://www.automationreadiness.eiu.com/](https://www.automationreadiness.eiu.com/)

Call for Comments: Artificial Intelligence (AI) Primer

[https://oecd-opsi.org/ai-consultation/](https://oecd-opsi.org/ai-consultation/)

Democracy in a Digital Society


Government Artificial Intelligence Readiness Index 2019


Industry 4.0 Opportunities Behind The Challenge Background Paper


OECD Principles on AI


Responsible use of artificial intelligence (AI) in Canada


World Economic Forum Inaugurates Global Councils to Restore Trust in Technology

References

1. https://www.epi.org/productivity-pay-gap/
4. “Elements of AI”
5. Metropolitan Government’s Online Procedures Enhancement for civil applications
This Background Paper was prepared collectively by members of the Working Group on “Trust and Public Debate in the Disinformation Age” steered by George Tilesch (Ipsos Global Affairs) ahead of the Policy Dialogue. It is intended as a starting point for the Action Lab discussions that will take place during the Dialogue and as a basis for the final “Call to Action”.

Working Group Members were Prof. Raja Chatila (EU High Level Expert Group on AI), President Laura Chinchilla (WLA-CdM), Nicholas Davis (WEF), Andres Delich (OEI), Clara Hanot (EU Disinfolab), Omar Hatamleh (International Space University), Prime Minister Yves Leterme (WLA-CdM), Maya Mirchandani (Observer Research Foundation), Susan Ness (Transatlantic Working Group) and Mark Surman and Sarah Watson (Mozilla Foundation). The ideas expressed in this Background Paper do not necessarily represent the views of individual Working Group members.
Introduction

Digital technologies have opened new channels of communication and coalition-building that allow for direct interaction between political leaders and citizens. They have also created a space for the expression of political ideas that might otherwise not find their way to the political debate. Digitalization has lowered the barriers for citizens to engage in nation-wide and global political conversations and amplified them. It is also one of the factors that has facilitated access to the political market for new actors, sometimes with a strong and often unscrupulous communications and digital strategy.

While serving as an alternative outlet for free speech in countries where this right is not a given, digital technologies also bring about new information challenges such as the extensive use of a persuasion/manipulation architecture that are derived from the current engagement model of social media that fuel the Attention Economy. Unwanted side effects of online anonymity, zero-cost publishing and content retransmission favor the propagation of political messages, hate speech, extremist and polarizing ideas. AI Algorithmic content curation, whereby AI recommendation engines, search engines and newsfeeds prioritize content based on each user’s profile, have created echo chambers that push online citizens further into ideological bubbles and away from multi-faceted analysis and block the construction of public debate. AI-driven technologies are designed to catch and capture our attention to maximize the time spend and attention to sell ads.

AI is no longer science fiction: it is now. While narrow in its focus at this point, it is already omnipresent, especially when it comes to rewiring our minds and the remnants of our Digital Commons. Powering all major digital social platforms that we are using, AI is now heavily influencing our information, conversations, choices as citizens and consumers, our relationships – our very lives. Therefore, we need to understand and analyze how AI technologies are being designed to exploit our human vulnerabilities as individuals and societies for economic and political gain – or pure disruption. The damage and transformation that has already been done with AI putting manipulation on steroids has fundamentally shaken the public dialogue and democratic prerequisites: the time to act is now.

AI is a double-edged sword, with both threats and boons to democracies worldwide. How can democratic societies harness this technological force of unprecedented power for good and use it to the public advantage? The imperative is not just fast adaptation of agile policymaking but also setting a shared vision of clear choices of what kind of AI is important to fuel the public dialogue and support democratic institutions. Policy and citizens should work to fulfill “Good AI’s” promise for inclusive societies and democratic guardians: an unforeseen level of potential public engagement. Good AI is designed and implemented based on the public’s values, as articulated through a deliberative and inclusive dialogue between experts and citizens – and powering such dialogue at the same time. This is what should drive policy’s and society’s expectations of new AI Tech towards AI Constructors. Chosen AI Solutions should be less “Artificial”, but bring “Augmented Intelligence”: a sustainable, humanistic vision that empowers both digital citizenship as well as radical speed and efficiency boosts for democratic governments to craft connected policy.

Trust Dismantled

Trust has been eroding globally for some time but the decline has accelerated and been further disrupted by digital waves of increasing intensity. We are at a precarious point in history when our fundamental institutions - government, public services, the media, corporations – are not seen as trustworthy by a majority around the world overall. The public perceives the most important drivers of trustworthiness to be reliability,
transparency and responsible behavior: there is clearly a gap as of now.

Digital communications and social media behavioral phenomena have aggravated the "perils of perception": perceptions on trust are often out of line with reality. For example, research worldwide on attitudes towards refugees shows that people tend to rely on personal experience to build their own "reality", and dismiss vertical information coming from governments, media, and intellectuals. With Millennials slowly taking charge, societal trust is significantly lower than with preceding generations²: low trust environments are fertile ground for disinformation as the public loses confidence in impartial arbiters of a common set of truths. This spirals further into understanding voters of populists who are primarily characterized by a very low level of trust in other people - populism is the political manifestation of the fear of having one's place in society threatened and of having lost control over one's life.

Journalists are now one of the Top 3 least trusted professions³. The media made several mistakes fighting its economic decline and the decimation of the journalist workforce: including the pressure of optimizing content for social media, failing ad-based business models, 24-hours "Breaking News" attention desperation, ethical decline and unbound partisanship. One-third of the public trusts media less than they did 5 years ago; more than 6 in 10 think that online news sources contain a ‘great deal’ or ‘fair amount’ of disinformation⁴.

Digitalization has led to the disappearance of online public dialogue and to the revelation that ‘connectedness disconnects.’ In the social media world, a concentrated, loud minority opinion of a few percent can create the illusion of being the aggressive majority, especially if that opinion is magnified by fake agents, both humans and bots. Both the disappearing middle and waning empathy are detrimental to democracy: we are being pushed into corners/camps due to a combination of exploitative tech (e.g. subpar AI Labeling of our views, recommendation engines keeping us in our echo chambers) and psychological weaknesses such as vilification of dissent or groupthink. Understanding and engaging the silent, confused, fearful “bystander” majority who accidentally handed the town square to extremists is key.

AI-driven disinformation

The last few years have brought the overuse of the term “fake news” - until it is devoid of any meaning, so we will use disinformation/misinformation consistently. The public’s awareness and frustration with online manipulation is clearly demonstrable. Four in five global citizens admit to having been exposed to misinformation, and four in ten (44%) admit to being duped by it. However, few can agree who should police the digital commons and determine what is fake⁵. The extreme challenge here is that disinformation often involves legal but manipulated content. Unlike terrorism, violent extremism, or child pornography -- universally considered unprotected speech -- platform action on disinformation is not as clear cut: understanding the distinction between legal and illegal speech is important.

Strong majorities support all forms of actions to resist online disinformation, save for extreme government censorship. 75% say that social media has too much power and platforms are responsible for distrust. Public opinion is inconsistent on how to deal with this matter: six in ten global citizens say that it is acceptable to temporarily cut off access to social media platforms during times of crisis to prevent the spread of misinformation. Equally, six out of ten also agree that temporarily cutting off social media platforms at times of crisis is unacceptable because social media is the primary source of news and information for many people⁶.
Propaganda and mass manipulation are far from new, but boosted with AI, they can be even more detrimental, to where they can threaten democratic collapse. In the context of democratic threats, AI algorithms provide unparalleled capabilities in mass social media data sourcing/scraping, processing and analyzing. AI’s ability to hyper-target misinformation on individuals and groups based on data we share unknowingly is already being exploited extensively. AI tech is used to impersonate real people with the ability to create mass fake profiles, show fake power and influence and sway public opinion. The impact is widespread and access is low-barrier: an officially retired powerful fake commenter AI engine that can imitate the style of any person reappeared in the public domain with an investment of a mere $5000.

The disinformation arsenal is increasing to higher levels of sophistication every day. With AI becoming core to digital manipulation, one emblematic weapon of disinformation is deepfake technology. Deepfakes are AI-manipulated videos created to look legitimate and can be AI-produced in myriad different versions, hyper-targeted at the individual: even current AI technology makes deepfake creation barriers very low. Fighting them will be a very resource-intensive race. Catching a deepfake only allows the AI network to produce a new one with better quality in a matter of seconds. The prevalence of deepfakes also enables the “liar’s dividend;” where a politician can claim that a damaging real video is a “deepfake,” further obliterating the meaning of evidence and factuality.

**AI: Algorithms for the Public Interest?**

There is a growing consensus regarding the identification of a wide variety of challenges and threats on the public dialogue and social cohesion by AI-powered technologies. AI-based hate speech detection on social media is reported to be racially biased. Recommendation engines and videos on autoplay are claimed to take citizens down a rabbit hole of radicalization. Data ignorance increases people’s vulnerability to AI-powered exploitation and mass manipulation and adds to their sense of powerlessness. Armies of underpaid human moderators are contracted to label data and help bridge the gap with AI’s weaknesses in detecting disinformation and hate speech: but will they override a machine’s decision? Basic legal frameworks are lacking when it comes to establishing distributed legal responsibility with disinformation campaigns; there are many actors in the digital information world, including tech providers, advertisers, the platform, the medium, moderators, and the user, and it is unclear who are liable.

Citizens worldwide react to these perceived threats with a mixture of confusion and concern: 40% globally are worried about AI use, with the concerned taking the lead: 1 in 5 even wants to ban AI outright. More respondents agree than disagree that governments’ and companies’ use of AI should be more strictly controlled. AI Algorithms we use every day are perceived to be biased by majority populations globally, especially in the developed world. Due to a lack of transparency, a perception that they are exploitative by design and the absence of a human element from decision-making are cited by naysayers.

While they are fulfilling a semblance of a “global public town square” function, many feel that AI-powered social media algorithms currently are not in line with the public interest. Policy’s struggles with understanding and regulating social media are far from being resolved (see Facebook US Senate and EU Parliament hearings). With legislature in limbo, AI pack leader companies stick to their own Terms of Use, lacking adequate regulatory frameworks: some of them say they self-regulate, some call for being regulated. Current challenges range through policing hate speech (e.g. Facebook’s alleged role in Myanmar genocide); data abuse combined with hyper-targeted (political) advertising (e.g. Cambridge Analytica case and
many more); the threat of surveillance capitalism and obscure data monetization; universality of values vs. different cultural norms; establishing accountability in misinformation campaigns; and general transparency and explicable of algorithmic decisions.

The test of our time, largely driven by the Algorithmic Economy, is how to increase both AI understanding, connectedness and consent of policymakers, experts, business and general public in concert? As a first step, the last few years were marked by simultaneous efforts to create consensual AI Ethics frameworks, from both top-down and bottom-up. Currently, there are 285 AI Ethics Code proclamations co-existing or competing. How can we integrate, simplify and make them policy-ready, especially the ones that affect public dialogue, which are most tangible for the general public?

New frontiers: Upgrading democracy for the AI Age

The dawn of the AI Age should be a stepping stone to deep reflection for policymakers on political philosophy and on remaking democracy. Anti-establishment sentiment and anti-political class exhaustion vs. perceived AI efficiency reached such levels that an EU study found that on average, 33% of European citizens would allow AI to make important decisions about running their country\(^2\). The public may be ill-informed about AI but their concerns need to be addressed properly.

The very essence of democracy is in turmoil: falling trust, growing empathy gap, tribalization, purposelessness, diluted responsibility, emotional reactions to complexity and abundant cognitive biases are all significant dangers. If adaptation is unguided and unmediated, chances are that our future democracies, while becoming more and more “direct”, will very much resemble the increasingly disturbing face of social media. For responsible citizens and policy actors worldwide, conscious or laissez-faire disruption can only be countered with deep understanding, modernized deliberation, connected policy and concerted action to rebuild the global town square for the 21st Century.

Forging the path ahead: Policy Recommendations

While the work seems enormous and sometimes isolated, there are many efforts underway – government initiatives, bills proposed, and multi-stakeholder or corporate initiatives to address artificial intelligence in the context of disinformation. A few examples:

- In the recent Christchurch Call during UNGA, platforms agreed to reorganize GIFCT, to better coordinate with governments on identifying, tagging, cataloguing and removing violent extremist and terrorist content.

- In 2018, the European Commission persuaded four online platforms and a few advertising industry trade associations to agree on a self-regulatory Code of Practice to address online disinformation. From the companies’ June 2019 reports, the EC found that progress was made on transparency of political advertising, “that actions taken against abusive use of bots and fake accounts have helped to detect, debunk and close down manipulation activities targeting the elections Google, Facebook and Twitter improved the scrutiny of ad placements to limit malicious click-baiting practices and reduce advertising revenues for purveyors of disinformation,
for instance by taking down ads and closing ad accounts due to deceptive or inauthentic behavior."

- There is ongoing discussion among platforms and civil society about creating a searchable database of disinformation and hacked content, in addition to sharing the technology for smaller platforms to be able to better detect disinformation and this dialogue needs to be encouraged.

1) Empowering Digital Citizens

Existing research on cognitive biases - especially tech-induced ones - should be guiding the hands of policymakers. While policy can set the rules, the final frontier is the human mind, so double resilience should be built up that pays equal attention to providing protection from digital manipulation and to building self-awareness to counter cognitive threats. To ensure a free, publicly vocal and informed citizenry, this should take shape as a "Live Digital Citizen Curriculum" with topics including Empathy, Critical Thinking, Media Literacy, Data Literacy, AI Literacy, and Security Literacy. This should be built into public education systems as well as pushed into adult education and mass education/communication programs.

2) Rewiring Media

To earn back trust, both traditional and social media needs to be in line with the public's expectations for trustworthiness: new exemplary online behaviors should originate from unbiased individual experts/organizations that hold majority society's trust and set new norms of responsible behavior in the global digital town square. Combined efforts are needed for government regulation and global ethics consensus, and for a fundamental rethinking of the technological and economic incentives that lead some media outlets to break ethics rules.

Factchecking Boost: Live, AI-powered factchecking should be a global multi-stakeholder effort with robust public investment and with the aim of integrating and elevating a multitude of isolated factchecking efforts worldwide. Despite public investment, direct government control should be minimized. Public interest AI development should prioritize finding a way to fact-check fast-breaking stories where rumors are rampant and substantiation thin, as this is essential in crisis situations. The signatories to the recent UNGA Christchurch Call have made the latter a priority, and there is also some movement in the EU regarding a sustainable European factchecking network.

3) Rebuilding the Core

While a lot of research has been going into dissecting populism and extremism lately, the focal point should be redirected at understanding the disillusioned majority and actively rebuilding the democratic core. For AI to serve the public dialogue we have to rethink rules (and regulatory requirements) to algorithmically reward trust, constructiveness and emphasize the spectrum of opinions instead of corners. With the speed of changes in public opinion, this can only be done efficiently with AI-powered citizen intelligence: real-time, multi-data source, public-facing citizen intelligence platforms that analyze and track the whole spectrum of public dialogue, with the aim of finding common denominators for policy formulation. AI can also be used to bridge to bridge quantitative reach and qualitative depth for mass citizen digital feedback and conversations. Boosting nascent, currently underfunded civic tech efforts to fostering public deliberation online via institutionalized processes is therefore key. These should be paired with a public interest fact base: shared, accessible, independently and reframed as a public service. Some core elements of how Wikipedia has been built and some of its functions may be considered as building blocks for the solution, but in a more dynamic/agile manner.

4) Public Interest Technology Stack

Rethink the digital town square: Multi-stakeholder efforts should establish Public Interest Tech Labs that produce public dialogue-enhancing solutions
(platforms, app, plugins etc.) with the speed of startups, have access to funding and create interoperable tech. Experiment with new “Public Interest Tech” category at that the international level that could be e.g.:

- Publicly funded, publicly accountable, nonprofit public dialogue platforms (e.g. similar to Signal - funded by a nonprofit foundation - vs. Facebook Messenger/WhatsApp in the messaging world)

- Joint venture between responsible Tech Platforms and governments to co-create a new breed of public benefit fora that is not ad-based with vetted information, citizens owning their data and algorithms being wired to fostering consensus, civility and harmony.

It is possible under the umbrella of the UN or another international organization: it should have public-private funding but be shielded from direct government influence (forbearers include the BBC, C-SPAN [entirely funded by the cable TV industry as a public service], or PBS).

- A regulatory requirement for Tech platforms to create “public benefit sandboxes” as part of their platforms for open public dialogue tech experimentation.

5) Good AI Policy for Public Dialogue

Aggregate AI Ethics Codes Using both AI and human curation methods, the next step should be to build an overarching global framework of actionable, policy-ready AI Ethics Codes that set a clear and actionable vision that can be adopted by global institutions.

Prioritize AI Dev Directions: Policy actors should feel emboldened to pick, steer and mandate certain beneficial AI development directions as well as to monitor and scrutinize high-risk ones. There are nascent technologies (e.g. Contextual AI) that prioritize symbiotic machine-human coexistence and collaboration: policy needs to pick those to reduce the AI Black Box\textsuperscript{14} effect in new AI regulations. Principled public investment into these technologies should start at the basic research level funding proactively, instead of being in constant follower mode due to diverging interests regarding AI usage priorities from business, intelligence services or the military.

“Open Sourcing” AI Policy frameworks. For lots of states, especially in the developing world, robust AI Policy knowledge is not accessible locally. Dedicated public interest AI Policy Centers should be set up regionally that have the combination of cutting-edge expertise and local familiarity to serve a multitude of countries. These centers could produce up-to-date AI Policy blueprints that can be localized with less effort, enabling the AI Ethics to AI Policy fast track as well as global harmonization.

6) Protecting Elections

“Election War Rooms” We need new structures that bring tech giants and governments together to protect elections from interference. US has started to adopt this thinking slowly. The concept of a “war room,” where tech companies, election authorities, cyber-security and intelligence agencies share information in order to disrupt domestic or foreign interference with elections should be made mandatory with some form of real-time access for the public and researchers, without benefitting bad actors.

7) Regulating Social Media

Current platforms oftentimes equate regulation with punitive action and this needs to change. Because of the heavy resource and skill needs of AI Technology, a new consensus should be reached and buy-in from AI platforms ensured. AI regulation should be pre-emptive rather than reactive through continuous connectedness to both the speed of AI technological evolution and shifting public opinion. Multi-stakeholder structures should be set up for to align interests and motivations for Public Interest AI development in a joint framework.

Mandatory transparency: Increase regulatory accountability of existing platforms towards citizens to serve. Social media platforms should be made transparent and mark/disclose
all amplified/automated content (e.g. bot activity, ads) in a way that is easy to interpret for the user. Explore the concept of a dedicated digital public dialogue regulatory agency that monitors in real time whether social media platforms are clear in their moderation standards (terms of use), fair in their conduct (opportunity for redress), and transparent. For example, a recent French government proposal (May 2019) would establish a regulatory regime based on transparency and accountability.

New Civil(ity) Code: Civil society, communities and companies should insist on greater civility on platforms - an area that should be emphasized and expanded. Current efforts are focused on AI monitoring and deleting of hate speech, but that is just the tip of the iceberg. To go beyond, we may consider to algorithmically encourage civility/constructiveness while staying very conscious at protecting free speech at the same time. A good instrument could be AI (Natural Language Processing) tone monitoring on social media and automatic up/downranking posts/comments accordingly: again, this will pose challenges vis-à-vis free speech, but if successfully mitigated, could set new online behavioral standards. Governments should support the creation of such a public benefit technology in an open source manner that even smaller platforms could build in. In concordance, platforms should be further encouraged to prohibit harassment in their terms of service/community standards, and to orchestrate greater cross-platform cooperation to track harassing behavior.

Trusted accounts: With platforms both new and old, algorithmically reward gaining trust from across the opinion/ideological spectrum. Trust needs to be earned by enduringly nonpartisan users in a long time, showcased with mandatory "trust score badges" on social media profiles. Additional trust-building features may be considered on social media, such as mandatory ID Verification when political activity exceeds a predetermined level (Airbnb) or expertise badges (Linkedin).

8) Regulating Adversarial digital campaigns

Foreign interference with elections merits a different set of responses, especially if using information operations. Governments should increase their vigilance, heavily regulate and limit state-backed mass manipulation campaigns: diplomacy, sanctions, blocking and retaliation are among the tools they can use. Some platforms are working to identify and block state-backed mass manipulation campaigns (e.g. Twitter re: Hong Kong) working on banning political hypertargeting use - this needs to be encouraged, monitored and incentivized. Legislation to prohibit advertising by state-backed entities should be considered, but complications come when distinguishing between organizations like PBS, which is partially funded by a government agency and the likes of Russia Today, which is entirely dependent on Russian government support. Related legislation has stalled, at least partially because some political parties have an interest to uphold the chaotic status quo. (The Honest Ads Act which requires identification of sponsors of political ads is blocked in US Senate, but some platforms have voluntarily pledged to following its requirements.)

Constant real-time monitoring and evaluation of new digital impersonation tech should be mandatory and clear limits should be set with the potential regulatory hard ban on harmful AI technologies and their producers. (Open AI's self-ban example is a good one on AI fake commenting tech.) Some platforms have publicly earmarked funding for improving artificial intelligence to detect deepfakes and Facebook has recently launched a community challenge for improving deepfake detection. As per above, there are early attempts to legislate (e.g. California mandates to disclose the use of bots; Senator Feinstein has introduced a federal law that bans the use of bots in political campaigns, but chances are very low
that it will pass under the current administration.) The risks of mitigating manipulative content is moving from major platforms onto smaller ones if the latter do not have the capacity to monitor.

References

1. Ipsos “Trust: The Truth?” Study September 2019
2. Ipsos MORI Millennial Myths and Realities
3. (Journalists 26%, Ministers 22%, Politicians Generally 19%) Ipsos MORI 2018 Veracity Index
4. Ipsos June 2019 Global Advisor survey: Trust in Media
5. CIGI-Ipsos Global Survey on Internet Security and Trust
6. Ipsos Global Advisor August 2019
9. World Economic Forum/Ipsos Global AI Perceptions study July 2019
10. CIGI Ipsos Study 2019 on Internet Security & Trust
12. IE CGC European Tech Insight 2019