Club de Madrid/Boston Global Forum

POLICY LAB
FUNDAMENTAL RIGHTS IN AI DIGITAL SOCIETIES: TOWARDS AN INTERNATIONAL ACCORD

Issues paper from sub-committee 1:Opportunities and threat for fundamental rights in AI & digital societies.

Dr Paul Twomey
Sub-Committee Leader

Artificial Intelligence (AI) is reshaping human experience in ways not visible to, nor fully apprehended by, the vast majority of the world's population. The explosion of AI is having a notable impact on our present rights and future opportunities, determining the decision-making processes that affect all in today's society.

Enormous technological change is occurring. It promises great benefits and poses insidious risks. The proportion of risks to benefits will become apparent, depending on the pioneers and creators of this technology, and, in particular, on the clarity of their and the political classes' vision of the common good.

This issue paper from the Sub-Committee commences with a discussion of the issues posed by how AI (and its interrelated Big Data) is used in the work place, the market for consumer and citizen behavior, and in the military.

Then the paper turns to the questions of competition issues and impact on human rights. Further some principles for government responses are outlined.

Fourthly multilateral governmental responses to date are briefly sketched.

The above issues were discussed at a videoconference among some of the members of the Subcommittee. The final section of this paper indicates the suggestions for the Policy Lab from this discussion.

The Issues

The use of automated decision making informed by algorithms is penetrating the modern workplace, and broader society, at a rapid rate. In ways not visible to, nor fully apprehended by, the vast majority of the population, algorithms are determining our present rights and future opportunities. To consider just one

aspect of everyday life, automobile transportation, these algorithms help us drive our cars, determine whether we can get a loan to buy them, decide which roads should be repaired, identify if we have broken the rules of the road and even determine whether we should be imprisoned if we have (see Angwin et al. 2016).

Benefits

Big data and AI can provide many benefits. They can assemble and consider more data points than humans can incorporate and often provide less biased or clearer outcomes than humans making decisions.

Examples include the prevention of medical errors to increasing productivity and reducing risks in the workplace. Even in the explicitly human function of the human resources department, machine learning can improve job descriptions and provide more "blind" recruitment processes, which can both increase the pool of qualified candidates and boost recruitment of non-conventional applicants.[1] Written well, algorithms can be more impartial and pick up patterns people may miss, in this and other applications.
Many commentators point to the productivity benefits of AI. For instance, analysis by Accenture of 12 developed economies indicates that AI could double annual economic growth rates in 2035: "The impact of AI technologies on business is projected to increase labor productivity by up to 40 percent and enable people to make more efficient use of their time" (Purdy and Daugherty 2016). The World Bank is exploring the benefits of AI for development and in uses from predicting migration patterns to reducing poverty.[2] Others identify farming, resource provision and health care as sectors in the developing economies that will benefit greatly from the application of AI (see Ovenden 2016).

Impact on Employment

Much has been made of the impact of AI and related robotics on jobs, especially since Carl Benedikt Frey and Michael A. Osborne's 2013 paper estimating that 47 percent of jobs in the United States were "at risk" of being automated in the next 20 years. Debate has ensued on the exact nature of this impact: the full or partial erosion of existing job tasks, the impacts across sectors and across developed, emerging and developing economies. Forecasting such effects is inherently difficult. But a recent summary from the McKinsey Global Institute reflects a midway analysis.

> Automation technologies including artificial intelligence and robotics will generate significant benefits for users, businesses, and economies, lifting productivity and economic growth. The extent to which these technologies displace workers will depend on the pace of their development and adoption, economic growth, and growth in demand for work. Even as it causes declines in some occupations, automation will change many more — 60 percent of occupations have at least 30 percent of constituent work activities that could be automated. It will also create new occupations that do not exist today, much as technologies of the past have done…
> Our scenarios across 46 countries suggest that between almost zero and one-third of work activities could be displaced by 2030, with a midpoint of 15 percent. The proportion varies widely across countries, with advanced economies more affected by automation than developing ones, reflecting higher wage rates and thus economic incentives to automate….

[1] See firms like Textio (https://www.textio.com/) and Pymetrics (https://www.pymetrics.com).
[2] See www.measuredev.org/.

> Even if there is enough work to ensure full employment by 2030, major transitions lie ahead that could match or even exceed the scale of historical shifts out of agriculture and manufacturing. Our scenarios suggest that by 2030, 75 million to 375 million workers (3 to 14 percent of the global workforce) will need to switch occupational categories. Moreover, all workers will need to adapt, as their occupations evolve alongside increasingly capable machines. (Manyika et al. 2017, vi)

Whatever the specifics, the results are clearly going to be very significant for G20 economies and their citizens. And, if the rate of adoption continues to outpace previous major technological adoptions,[3] the scale of social dislocation is likely to be greater — which provides even more reason for the G20 to work now on a framework for AI adoption.

Risk of Bias

Code is written by humans and its complexity can accentuate the flaws humans naturally bring to any task.

Bias in the writing of algorithms, as a product of human endeavour, is inevitable, and can have chilling effects on individual rights, choices and the application of worker and consumer protections. Algorithms incorporate built-in values and serve business models, which may lead to unintended biases, discrimination or economic harm.[4] Compounding this problem is the fact that algorithms are often written by relatively inexperienced programmers who may not have a correct picture of the entire application or a broad experience of a complex world. The dependency of the workplace on algorithms imparts tremendous power to those who write them. These programmers may not even be aware of this power or the potential harm that an incorrectly coded algorithm could do. Researchers have discovered bias in the algorithms for systems used for university admissions, human resources, credit ratings, banking, child support systems, social security systems and more. Because the complex market of interacting algorithms continues to evolve, it is also likely that existing algorithms that may have been innocuous yesterday will have significant impact tomorrow.

AI is subject to two significant types of bias:

- bias in its coding (both in design and development), or
- selection bias in or distortion/corruption of its data inputs.

Either type can result in significantly flawed results delivered under the patina of “independent” automated decision making.

The Criticality of Truly Applicable and Accurate Data Inputs

While much contemporary commentary has focused on the question of bias, the long experience of software development teaches that the proper scope, understanding and accuracy of data have dominant impacts on the efficacy of programming. In simple terms, “garbage in, garbage out.” This relationship is

[3] See discussion in Lohr (2017).

[4] For instance, media reports (see, for example, Wexler 2017) have pointed out clear racial bias resulting from reliance on sentencing algorithms used by many US courts.

particularly true with AI. AI is a process of machine learning — or, more accurately, machine teaching. The inaccuracies in data often come from reflections of human biases or human judgments about what data sets tell us. The establishment of training data and training features is at the heart of AI. As Rahul Barghava (2017) says, **"**In machine learning, the questions that matter are 'what is the textbook' and 'who is the teacher.' "The more scrutiny these can receive, the more likely that the data will be fit for purpose. To consider one example, some local governments in the United States have been making more use of algorithmic tools to guide responses to potential cases of children at risk. Some of the best implementations involve widespread academic and community scrutiny on their purpose, process and data. The evidence is that these systems can be more comprehensive and objective than the different biases people display when making high-stress screenings. But even then, the data accuracy problem emerges: "It is a conundrum. All of the data on which the algorithm is based is biased. Black children are, relatively speaking, over-surveilled in our systems, and white children are under-surveilled. Who we investigate is not a function of who abuses. It's a function of who gets reported."[5] Sometimes the data is just flawed. But the more scrutiny it receives,

the better it is understood. In the workplace, workers often have the customer and workflow experience to help identify such data accuracy challenges.

Acceptance of data inputs to AI in the workplace is not just a question of ensuring accuracy and fit for purpose. It is also one of transparency and proportionality.

The crisis surrounding Facebook, over Cambridge Analytica's illicit procurement of millions of its users' private data to inform data-targeting strategies in the 2016 US presidential election, has shown that there is a crisis in ethics and public acceptance in the data collection companies. Among the many issues raised by that scandal, a subset includes:

- a realization of the massive collection of data beyond the comprehension of the ordinary user;
- the corporate capacity to collate internal and external data and analyze it to achieve personally recognizable data profiles of users, which the users neither knew about nor explicitly approved;
- the collecting of people's data without any contractual or other authority to do so; and
- the lack of transparency in the data collection processes, sources, detail, purposes and use.

These issues are more urgent when they have a direct impact on people's working lives. It is important, to meet the pressing needs of data accuracy and worker confidence, that employees and contractors have access to the data being collected for enterprise AI, and, in particular, for workplace AI. Data quality improves when many eyes have it under scrutiny. Furthermore, to preserve their workplace morale, workers need to be sure that their own personal information is being treated with respect and in accordance with laws on privacy and labour rights.

Including Community Interests

The present discussion about the ethics of data gathering and algorithmic decision making has focused on the rights of individuals. The principles for the adoption of AI need to include an expression of the policy concerns of the community as a whole, as well as those of individuals. For instance, the individual right of intellectual property protection may need to be traded off against the community interest in non-discrimination (which is also an individual's human right) and, hence, a requirement for greater

[5] Erin Dalton, deputy director of Allegheny County's Department of Human Services, quoted in Hurley (2018).

transparency as to the purpose, as well as the inputs and outputs, of a particular algorithmic decision-making tool.

Risk of Further Marginalization of the Vulnerable

AI, at its heart, is a system of probability analysis for presenting predictions about certain possible outcomes. Whatever the use of different tools for probability analysis, the problem of outliers remains. In a world run by algorithms, the outlier problem has real human costs. A society-level analysis of the impact of big data and AI shows that their tendency toward profiling and limited-proof decisions results in the further marginalization of the poor, the Indigenous and the vulnerable (see Obar and McPhail 2018).

One account reported by Virgina Eubanks (2018, 11) explains how interrelated systems reinforce discrimination and can narrow life opportunities for the poor and the marginalized:

> What I found was stunning. Across the country, poor and working-class people are targeted by new tools of digital poverty management and face life-threatening consequences as a result. Automated eligibility systems discourage them from claiming public resources that they need to survive and thrive. Complex integrated databases collect their most personal information, with few safeguards for privacy or data security, while offering almost nothing in return. Predictive models and algorithms tag them as risky investments and problematic parents. Vast complexes of social service, law enforcement, and neighborhood surveillance make their every move visible and offer up their behavior for government, commercial, and public scrutiny.

This excerpt highlights the issue of unintended consequences, particularly costly when they impact the marginalized. It is unlikely that the code-writers of the systems described above started off with the goal "let's make life more difficult for the poor." However, by not appreciating the power of the outcome of the semi-random integration of systems — each system narrowly incented by the desired outcomes for the common and the privileged — that is exactly what these programmers did.

The same concerns apply to the workplace. As one example, at first glance it may appear intuitive to record how far an applicant lives from the workplace for an algorithm designed to determine more likely long-term employees. But this data inherently discriminates against poorer applicants dependent on cheaper housing and public transport. As another, AI written around a narrow definition of completed output per hour may end up discriminating against slower older employees, whose experience is not reflected in the software model.

Over the past few decades, many employers have adopted corporate social responsibilities, partly in the recognition that their contribution to society is more than just profitability. As the AI revolution continues, it is essential that a concerted effort be made to ensure that broader societal responsibilities are not unwittingly eroded through the invisible operation of narrowly written deterministic algorithms that reinforce each other inside and beyond the enterprise.

Big data and AI should not result in some sort of poorly understood, interlinked algorithmic Benthamism, where the minority is left with diminished life opportunities and further constrained autonomy.

Humans Are Accountable for AI

There is a tendency by some to view AI, because of its complex and opaque decision making, as being separate from other products made by humans, and a unified entity unto itself. Such a notion is a grave error and one that fails to understand the true role of the human within the algorithm. It is essential to emphasize the human agency within the building, populating and interpretation of the algorithm. Humans need to be held accountable for the product of algorithmic decision making. As Lorena Jaume-Palasí and Matthias Spielkamp (2017, 6-7) state:

> The results of algorithmic processes…are patterns identified by means of induction. They are nothing more than statements of probability. The patterns identified do not themselves constitute a conclusive judgment or an intention. All that patterns do is suggest a particular (human) interpretation and the decisions that follow on logically from that interpretation. It therefore seems inappropriate to speak of "machine agency", of machines as subjects capable of bearing "causal responsibility"...While it is true that preliminary automated decisions can be made by means of algorithmic processes (regarding the ranking of postings that appear on a person's Facebook timeline, for example), these decisions are the result of a combination of the intentions of the various actors who (co-)design the algorithmic processes involved: the designer of the personalization algorithm, the data scientist who trains the algorithm with specific data only and continues to co-design it as it develops further and, not least, the individual toward whom this personalization algorithm is directed and to whom it is adapted. All these actors have an influence on the algorithmic process. Attributing causal responsibility to an automated procedure — even in the case of more complex algorithms — is to fail to appreciate how significant the contextual entanglement is between an algorithm and those who co-shape it.

A Human-centric Model Is Essential for Acceptance of AI and to Ensure a Safe AI Future

Hundreds of technical and scientific leaders have warned of the risk of integrated networks of AI superseding human controls unless governments intervene to ensure human control is mandated in AI development. The British physicist Stephen Hawking spoke of the importance of regulating AI: "Unless we learn how to prepare for, and avoid, the potential risks, AI could be the worst event in the history of our civilization. It brings dangers, like powerful autonomous weapons, or new ways for the few to oppress the many" (quoted in Clifford 2017); further, he warned, "it would take off on its own, and re-design itself at an ever increasing rate. Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded" (quoted by Cellan-Jones 2014).

More specifically within the workplace, big data and AI could result in a new caste system imposed on people by systems determining and limiting their opportunities or choices in the name of the code-writers' assumptions about the best outcome for the managerial purpose. One can imagine an AI-controlled recruitment environment where the freedom of the person to radically change careers is punished by algorithms only rewarding commonly accepted traits as being suitable for positions.

AI should not be allowed to diminish the ability of people to exercise autonomy in their working lives and in determining the projection of their own life paths. This autonomy is an essential part of what makes us human. As UNI Global Union (2018, 9) says, in the deployment of these technologies, workplaces should "show respect for human dignity [and] privacy and the protection of personal data should be safeguarded

in the processing of personal data for employment purposes, notably to allow for the free development of the employee's personality as well as for possibilities of individual and social relationships in the work place."

Microsoft (2018, 136) has called for a "human-centered approach" to AI. This approach is important not only to control AI's potential power, but to ensure — particularly in the workplace, including the gig economy — that AI serves the values and rights humans have developed as individuals in societies over the last centuries.

As *The Economist* (2018, 13) has concluded: "The march of AI into the workplace calls for trade-offs between privacy and performance. A fairer, more productive workforce is a prize worth having, but not if it shackles and dehumanises employees. Striking a balance will require thought, a willingness for both employers and employees to adapt, and a strong dose of humanity."

The Need to temper AI and related Big Data Manipulation of Users and Citizens

A long standing tenet of public policy in both advanced and emerging economies is that where an economic actor is in a position to manipulate a consumer – in a position to exploit the relative vulnerabilities or weaknesses of a person in order to usurp their decision making– society requires their interests to be aligned and punishes acts that are seen as out of alignment of the interests of the person. Individuals in some relationships, for example between priests-parishioners, lawyers-clients, doctors-patients, teachers-students, therapists-patients, etc., are vulnerable to manipulation through the intimate data collected by the dominant actor, and these types of relationships are governed such that the potential manipulator is expected to act in accordance with the interests of the vulnerable party. We regularly govern manipulation that undermines choice, such as when negotiating contracts under duress or undue influence, or when contractors act in bad faith, opportunistically, or unconscionably. The laws in most countries void such contracts.

When manipulation works, the target's decision making is usurped to pursue the interests of the manipulator; and the tactic is never known by the target. Some commentators rightly compare manipulation to coercion (Susser, Roessler, and Nissenbaum 2019). For coercion, a target's interests are overtly overridden by force and the target knows about the threat and coercion. Manipulation, on the other hand, overrides a target's choice subversively. Both seek to overtake the authentic choice of the target and just choose different tactics. In this way, manipulation has the goals of coercion and the deception of fraud. And offline, we regulate manipulation similar to the way we regulate coercion and fraud: to protect consumer choice-as-consent and preserve the autonomy of the individual.

Online actors, such as data aggregators, data brokers, and ad networks, can not only predict what we want and how badly we need it but can also leverage knowledge about when an individual is vulnerable to making decisions in the interest of the firm. Recent advances in hyper-targeted marketing allows firms to generate leads, tailor search results, place content, and develop advertising based on a detailed picture of their target. Aggregated data on individuals' concerns, dreams, contacts, locations, and behaviors allows marketers to predict what consumers should want and how to best sell to them. It allows firms to predict moods, personality, stress levels, health issues, etc. – and potentially use that information to undermine

the decisions of consumers. In fact, Facebook recently offered advertisers the ability to targets teens when they are 'psychologically vulnerable.'

All this information asymmetry between users and data aggregators has sky-rocketed in recent years.

The data collection industry is not new. Data brokers like Acxiom and ChoicePoint have been aggregating consumers' addresses, phone numbers, buying habits and more from offline sources and selling them to advertisers and political parties for decades. But the Internet has transformed the space. The scope and intimacy of the data collection and the purposes for which it is sold and used is rarely comprehended by users.

One reason for this is that much of the data is collected in a non-transparent way and mostly in a manner that people would not consider covered by contractual relationships. Many Internet users, at least in developed countries, have some understanding that the search engines and the e-commerce engines collect data on what sites they have visited and that this data is used to help tailor advertising to them. But most have little idea of just how extensive this commercial surveillance is. A recent analysis of the terms and conditions of the big US platforms shows that they collect 490 different types of data on each user.[6]

A recent study of 1 million web sites showed that nearly all of them allow third party web trackers and cookies to collect user information to track page usage, purchase amounts, browsing habits, etc. Trackers send personally identifiable information such as user's name, address, and email and spending details. These latter allow the data aggregators to then de-anonymize much of the data they collect (Englehardt and Narayanan 2016, Libert, 2015).

But cookies are only one of the mechanism used to collect data on people. Both little known data aggregators and the big platforms draw huge amounts of information from cell towers, the use of the devices themselves, many of the third party apps running on the user's device, Wi-Fi access, as well as public data sources and third party data brokers.

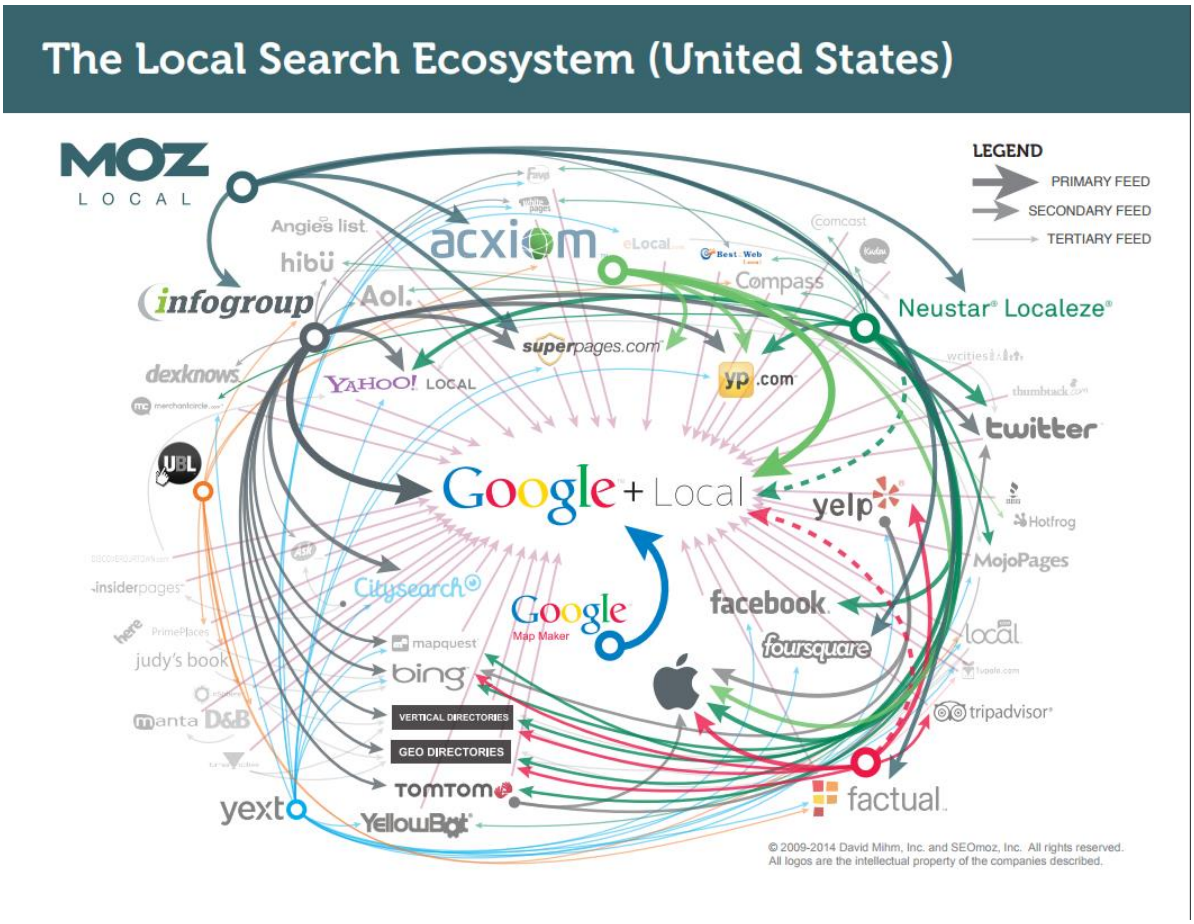
As the New York Times recently reported:

> Every minute of every day, everywhere on the planet, dozens of companies — largely unregulated, little scrutinized — are logging the movements of tens of millions of people with mobile phones and storing the information in gigantic data files. The Times Privacy Project obtained one such file [which] holds more than 50 billion location pings from the phones of more than 12 million Americans as they moved through several major cities... Each piece of information in this file represents the precise location of a single smartphone over a period of several months…It originated from a location data company, one of dozens quietly collecting precise movements using software slipped onto mobile phone apps.[7]

[6] See the publicly available data at https://mappingdataflows.com/

[7] "One nation, tracked An investigation into the smartphone tracking industry from Times Opinion" https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html?searchResultPosition=8

An indication of the scale and complexity of the collection and transfer of user data among web sites can be gleaned from the following diagram. Devised by David Mihm, a noted expert on search engine optimization, it shows the data feeds contributing to the US online local search ecosystem.[8]



It is data collection networks and markets like these, invisible to the vast majority of the people whose personal data is being collected, which enable Cambridge Analytica (of the 2016 US Presidential election fame) to claim that it holds to have up to five thousand data points on every adult in the US.[9]

AI in the military

[8] https://whitespark.ca/blog/understanding-2017-u-s-local-search-ecosystem/

[9] See "MPs grill data boss on election influence", 27 February 2018 http://www.bbc.com/news/technology-43211896

In 2015, a group of leading AI researchers and investors signed an open letter warning of the dangers of autonomous weapons. "The key question for humanity today is whether to start a global AI arms race or to prevent it from starting. If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable."[10] Today, many nations are pushing to apply AI for military advantage. While the phrase "AI arms race" is misleading – AI is a general technology enabler rather than a weapons system in itself – the rush to deploy it brings with it real risks. As Paul Scharre has written, "The widespread adoption of military AI could cause warfare to evolve in a manner that leads to less human control and to warfare becoming faster, more violent, and more challenging in terms of being able to manage escalation and bring a war to an end. Additionally, perceptions of a "race" to field AI systems before competitors do could cause nations to cut corners on testing, leading to the deployment of unsafe AI systems that are at risk of accidents that could cause unintended escalation or destruction."[11]

Partly in response to at least some of these concerns, the US Department of Defense adopted in 2020 a set of AI ethical principles encompassing five major areas:

1. Responsible. DoD personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities.
2. Equitable. The Department will take deliberate steps to minimize unintended bias in AI capabilities.
3. Traceable. The Department's AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedure and documentation.
4. Reliable. The Department's AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across their entire life-cycles.
5. Governable. The Department will design and engineer AI capabilities to fulfill their intended functions while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behavior.[12]

But when the issue of limiting the development of autonomous weapons systems has arisen for international discussion, there has been no consensus on legal action to limit their use. Such a limitation could be achieved through a new protocol to the Convention on Conventional Weapons (CCW), which has discussing this concern since 2014. While most states have recognised the need to retain some form of human control over these weapons, neither the United States nor Russia is willing to enter yet into negotiations of a limitations agreement.

---

[10] "Autonomous Weapons: An Open Letter from AI & Robotics Researchers," Future of Life Institute, 2015, https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1.

[11] Paul Scharre, "Debunking the AI race myth", *Texas National Security Review*, Volume 4, Issue 3, (Summer 2021) pp 121-132, at p 122.

[12] https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/

2 Impact on Rights

Protection against discrimination[13]

Our colleague Dr. Jutta Treviranus has written that one area that requires further emphasis is the tendency for machine learning to amplify, accelerate and automate existing discrimination against minorities and outliers. This inevitably occurs even when there is no bad intent or bias on the part of the developers or implementers. It is a diffuse effect that exponentially increases with each iteration and with each machine learning training session. This is not addressed by ensuring proportional representation and closing data gaps. It is not addressed by current AI ethics auditing tools that compare treatment of a specific protected identity group with treatment of the population as a whole,[14] or elimination of obvious human bias. The problem pre-dates AI, and big data analytics. The origin is in the principle of majority rules, evidence based on statistical reasoning and decisions based on probability. Prior to big data and AI there was a greater possibility of the determination of reasonable exceptions. Automated decisions remove this possibility and current AI ethics measures only make it harder to argue mistreatment. The impact is exponentially accelerated and amplified.

One group that feels the impact most is people with disabilities.[15] The only common data footprint of disability is sufficient difference from the mean that systems do not address your needs. Disability is at the outer edge of every justice seeking group, at the same time it is the world's largest minority.[i16]It has no bounded definition and because of stigma associated with disability, people often do not self-identify. Many people with disabilities represent an 'n' of one. People with disabilities tend to be the extreme small minorities and outliers in any population data set. The impacts of disability complexly compound through all aspects of life, including poverty, education, work, health, digital inclusion, etc. Scientific evidence and determinants of scientific "truth" for the general population tends to be wrong if you have a disability. Decisions guided by data will likely rule against you. The extreme injustice of this is that the more critical your needs, and the better off the general population is, the more likely your needs will not be addressed. The trivial needs of the many will overpower the critical needs of the few in majority-rules decision systems. The accelerating drift of data sets means the situation worsens as AI becomes more pervasive.

People with disabilities are most vulnerable to data abuse and misuse and current privacy protections do not work. Because of the uniqueness of the data footprint, people with disabilities can be easily reidentified with any aggregation of data. At the same time differential privacy removes the unique data needed to serve the needs of someone with a disability.

---

[13] My thanks to Jutta Treviranus for this section on discrimination.

[14] https://www.brookings.edu/research/auditing-employment-algorithms-for-discrimination/

[15] Treviranus, J., Gupta, A., (2020). Inclusive Designed Artificial Intelligence. In Schaffers, H. Vartianen, M., Bus, J., *Digital Innovation and Societal Change*. River Publishers, London, UK.

[16] Trewin, S., Basson, S., Muller, M., Branham, S., Treviranus, J., Gruen, D., Hebert, D., Lyckowski, N. and Manser, E., 2019. Considerations for AI fairness for people with disabilities. AI Matters, 5(3), pp.40-63.

There is a discriminatory hierarchy even within disability. The more you are at the margins, the harder it is to use personalization systems intended to meet your needs. Every learning model tends toward a mean. To train the model to serve your unique needs requires a great deal of "swimming upstream." Thus "struggling students" are less likely to be served by instructional tutors.[17] Individuals whose speech differs greatly from the norm have greater difficulty using speech recognition systems and people who are blind but live in greater poverty will not be able to use pattern recognition systems intended to replace vision.

(The beneficial examples listed in the paper do not address the bias against disability in hiring systems. Behavioural science has many of the built-in biases of other data analytics systems. Even if the assessments are made accessible in systems such as "Pymetrics," the accessible systems will be more complex and involve more cognitive load.)

The only means to address the discriminatory drift is to support bottom-up data with user-controlled post-hoc aggregation.[18] To protect privacy, intelligence would be on the personal device. Individuals and communities would co-create the data applications. Data aggregation could be stewarded by cooperative data trusts.[19]

A compromise measure would be mandatory periodic data refresh, to fight both dataset pollution and the discriminatory drift. The application of AI to assist in decisions or filter individuals should require a disability impact assessment. A disability impact assessment will be an indicator of treatment of all forms of difference. Data nutrition labels[20] and systems that signal when AI guidance is likely not to apply within a given decision instance would help to reduce the risk of false positive security decisions or critical decisions in health or education.

The current shortcomings of even ethical AI, namely difficulty addressing the unexpected, and difficulty transferring to new contexts, would benefit from more diversity-supportive decision systems. The greatest diversity is at the margins of a data set. Thus, addressing the barriers and discrimination against people with disabilities would be beneficial to the entire population.

Competition Issues

Barriers to entry exist in many AI driven digital markets due to network effects (the value of services rise with the number of users), first mover and scale advantages in accumulating machine learning training data sets, and a range of other factors. The treatment of the resulting power asymmetries should be treated

[17] Treviranus, J. (2021). Learning to Learn Differently. In: Holmes, W. (ed) *Ethics of AIED. Who Cares?*" Taylor & Francis Group, Oxon, UK.
[18] Tchernavskij, P. (2019). Designing and Programming Malleable Software (Doctoral dissertation, Université Paris-Saclay (ComUE)).
[19] https://ec.europa.eu/jrc/communities/sites/jrccties/files/micheli.pdf
[20] https://datanutrition.org/

more analogously to the regulation of natural monopolies offline.[21] Many jurisdictions, including the EU (via the Digital Services Act), have begun the process of investigations and accompanying legislative reform to re-establish the conditions for effective competition in these markets.

Many have also promoted the collection of large data sets under public commons rules to achieve a number of objectives:

- restore agency to people regarding how their data is circulated and used – recognising that people often have a range of desires that include but are not limited to the commercial sphere,
- help to increase the amount and quality of data potentially available to all firms, not just the largest technology platforms, to build a more level playing field, boosting competition in line with competitiveness goals, values and fundamental rights.

One country where authorities are moving to reduce big platform/AI players' market power is China. The ongoing shake-up of the Tech sector is partly driven by a sense that big data sets should be more freely available to smaller players in the market, as well as competing state owned enterprises. As *The Economist* noted on 14 August 2021:

> As a guiding principle, the vice-premier, Liu He, recently stated that China is moving into a new phase of development that prioritises social fairness and national security, not the growth-at-all-costs mentality of the past 30 years…
>
> Start with data. Europe and some American states, such as California, have devised laws that seek to protect consumers from the misuse of their personal information by large companies. China has put similar rules in place; in some cases they are more severe than in the West. But Chinese regulators are going further. In a largely ignored, jargon-filled policy paper from the State Council, China's cabinet, in April last year, data were named as a "factor of production" alongside capital, labour, land and technology…
>
> China's new data policy remains a work in progress. The Data Security Law will come into force on September 1st and the Personal Information Protection Law is due to be adopted by China's rubber-stamp parliament soon. It is unclear how they will be enforced, though data specialists intuit that many types of data currently held by internet giants could eventually be traded on government-backed and private exchanges. Ant, for example, is already being prodded by authorities to open up its vast stores of personal financial data to state-owned companies and smaller tech rivals.[22]

---

21 For a summary of the literature on the regulation of natural monopolies, see Joskow (2007). For a recent analysis, see Ducci (2020).

22 "What Tech does China want", *The Economist*, August 14, 2021. https://www.economist.com/business/what-tech-does-china-want/21803410

AI and Human Rights

Much of the public critique of AI focuses on racial or sex bias in learning data sets and or the outputs of the systems. But the right to non-discrimination is not the only human right vulnerable to AI deployments. Indeed broadly deployed AI affects nearly every internationally recognized human right, from the rights to privacy and freedom of expression, to the rights to health and education.[23]

Accessnow, an international NGO focused on the digital rights of users at risk around the world, has issued a review of how human rights are challenged by AI and makes the following broad recommendations:

1. Comprehensive data protection legislation can anticipate and mitigate many of the human rights risks posed by AI. However, because it is specific to data, additional measures are also necessary.
2. Government use of AI should be governed by a high standard, including open procurement standards, human rights impact assessments, full transparency, and explainability and accountability processes.
3. Given the private sector's duty to respect and uphold human rights, companies should go beyond establishing internal ethics policies and develop transparency, explainability, and accountability processes.
4. Significantly more research should be conducted into the potential human rights harms of AI systems and investment should be made in creating structures to respond to these risks.[24]

Much detailed discussion of the effect of AI on human rights (at least as they have evolved in a European context) is contained in the work of the Ad Hoc Committee on Artificial Intelligence at the Council of Europe[25] (CAHAI), which has in its Feasibility Study[26] addressed both the need for international regulation, as well as the instruments through which this could be achieved - identifying a need for a combination of binding and non-binding instruments, some with horizontal character, some vertical or sectorial.

Some Framework Principles

Over the last several years, the author has developed (in response to the G20 policy processes) a series of principles which give a framework for how governments could seek to protect the rights and well-being of citizens and workers in AI infused world. Partly to make the principles more palatable across a range of political systems, the author has cast many of the principles mostly in the context of the future of work (rather than broader civic life). Every government, not just liberal democracies, is confronted with the challenge of AI in the workplace and the need to maintain worker confidence.

---

[23] The rights being referred to are protected in the three main treaties: the Universal Declaration of Human Rights (UDHR), the International Covenant on Civil and Political Rights (ICCPR), and the International Covenant on Economic, Social and Cultural Rights (ICESCR).

[24] See *Human Rights In The Age Of Artificial Intelligence* https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights

[25] See https://www.coe.int/en/web/artificial-intelligence/home

[26] See https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da

The first set of seven framework principles relates to the collection of data in the work environment.

Right to know data is being collected, for what and from where: Workers, be they employees or contractors, or prospective employees and contractors, must have the right to know what data is being collected on them by their employers, for what purpose and from what sources.

Right to ensure worker data is accurate and compliant with legal rights to privacy: An important feature for worker understanding and productivity is to ensure that workers, ex-workers and job applicants have access to the data held on them in the workplace or have the means to ensure that the data is accurate and can be rectified, blocked or erased if it is inaccurate or breaches legally established rights to privacy. The collection and processing of biometric data and other personally identifiable information (PII) must be proportional to its stated purpose, based on scientifically recognized methods, and held and transmitted very securely.

Principle of proportionality: The data collected on present or prospective employees or contractors should be proportional to its purpose. As one group has proposed: "Collect data and only the right data for the right purposes and only the right purposes, to be used by the right people and only the right people and for the appropriate amount of time and only the appropriate amount of time."

Principle of anonymization: Data should be anonymized where possible. Data with PII should only be available where it is important to the data collection's prime purpose, and its visibility must be limited to the employee and the relevant manager. Aggregated, anonymized data is preferable for many management and productivity purposes.

Right to be informed about the use of data: Employees and contractors should be fully informed when either internal or external data (or both) has been used in a decision affecting their career. Any data processing of present or prospective employees' or contractors' data should be transparent and the available for their review. The right to understand and appeal against both the rationale employed and the data used to achieve that rationale is essential to safeguard present or prospective workers against poor or inaccurate input data or discriminative decisions.

Limits to monitoring of the workplace by employers: Proportional data collection and processing should not be allowed to develop into broad-scale monitoring of employees or contractors. While monitoring can be an indirect consequence of steps taken to protect production, health and safety or to ensure the efficient running of an organization, continuous general monitoring of workers should not be the primary intent of the deployment of workplace technology. Given the potential in the use of such technology to violate the rights and freedoms of the persons concerned, employers must be actively engaged to ensure that the use is constrained to specific positive purposes, so as not to breach these rights. This principle is not only a matter of workplace freedoms, but also a practical step toward maintaining morale and productivity.

Accuracy of data inputs and the "many eyes" principle: Employers should ensure the accuracy, both in detail and its intended purpose, of the data models and sources for AI. Poor data results in flawed decision

making. Training data and training features should be reviewed by many eyes to identify possible flaws and to counter the "garbage in, garbage out" trap. There should be a clear and testable explanation of the type and purpose of the data being sourced. Workers and contractors with experience of the work processes and data environment of the firm should be incorporated into the review of data sources. Such data should be regularly reviewed for accuracy and fit for purpose. Algorithms used by firms to hire, fire and promote should be regularly reviewed for data integrity, bias and unintended consequences.

An additional seven principles focus on AI in the workplace.
Focus on humans: focus: Human control of AI should be mandatory and testable by regulators.
AI should be developed with a focus on the human consequences as well as the economic benefits. A human impact review should be part of the AI development process, and a workplace plan for managing disruption and transitions should be part of the deployment process. Ongoing training in the workplace should be reinforced to help workers adapt. Governments should plan for transition support as jobs disappear or are significantly changed.

Shared benefits: AI should benefit as many people as possible. Access to AI technologies should be open to all countries. The wealth created by AI should benefit workers and society as a whole as well as the innovators.

Fairness and inclusion: AI systems should make the same recommendations for everyone with similar characteristics or qualifications. Employers should be required to test AI in the workplace on a regular basis to ensure that the system is built for purpose and is not harmfully influenced by bias of any kind — gender, race, sexual orientation, age, religion, income, family status and so on. AI should adopt inclusive design efforts to anticipate any potential deployment issues that could unintentionally exclude people. Workplace AI should be tested to ensure that it does not discriminate against vulnerable individuals or communities.
Governments should review the impact of workplace, governmental and social AI on the opportunities and rights of poor people, Indigenous peoples and vulnerable members of society. In particular, the impact of overlapping AI systems toward profiling and marginalization should be identified and countered.

Reliability: AI should be designed within explicit operational requirements and undergo exhaustive testing to ensure that it responds safely to unanticipated situations and does not evolve in unexpected ways. Human control is essential. People-inclusive processes should be followed when workplaces are considering how and when AI systems are deployed.

Privacy and security: Big data collection and AI must comply with laws that regulate privacy and data collection, use and storage. AI data and algorithms must be protected against theft, and employers or AI providers need to inform employees, customers and partners of any breach of information, in particular PII, as soon as possible.

Transparency: As AI increasingly changes the nature of work, workers, customers and vendors need to have information about how AI systems operate so that they can understand how decisions are made. Their involvement will help to identify potential bias, errors and unintended outcomes. Transparency is not necessarily nor only a question of open-source code. While in some circumstances open-source code will

be helpful, what is more important are clear, complete and testable explanations of what the system is doing and why.
Intellectual property, and sometimes even cyber security, is rewarded by a lack of transparency. Innovation generally, including in algorithms, is a value that should be encouraged. How, then, are these competing values to be balanced?
One possibility is to require algorithmic verifiability rather than full algorithmic disclosure. Algorithmic verifiability would require companies to disclose not the actual code driving the algorithm but information allowing the *effect* of their algorithms to be independently assessed. In the absence of transparency regarding their algorithms' purpose and actual effect, it is impossible to ensure that competition, labour, workplace safety, privacy and liability laws are being upheld.[27]

When accidents occur, the AI and related data will need to be transparent and accountable to an accident investigator, so that the process that led to the accident can be understood.

A related principle is data governance of record keeping: Long term data governance throughout the AI system lifecycle should be required to ensure that data used in AI systems is accurate, complete and appropriate and is stored in a safe and secured environment. Further appropriate records of the data management methodologies should be maintained.

Accountability: People and corporations who design and deploy AI systems must be accountable for how their systems are designed and operated. The development of AI must be responsible, safe and useful. AI must maintain the legal status of tools, and legal persons need to retain control over, and responsibility for, these tools at all times.

Workers, job applicants and ex-workers must also have the "right of explanation" when AI systems are used in human-resource procedures, such as recruitment, promotion or dismissal.[28] They should also be able to appeal decisions by AI and have them reviewed by a human.

Sustainability. AI should be able to detect unintended environmental harm and automatically disengage if it occurs, or allow deactivation by a human. It is particularly important that AI and autonomous devices deployed in agriculture and mining should be designed and monitored for long term environmental sustainability and maintenance of biodiversity. Leaving agriculture just in the thrall of the efficiency motive will result in monocultures and loss of food diversity.

Principles to protect the citizen as consumer as well as worker

In the offline world, we have developed safeguards to ensure that those with intimate knowledge of others do not exploit vulnerabilities and weaknesses of individuals through manipulation. Yet, online data

[27] This is explored to some degree by the Global Commission for Internet Governance (2016, 45).

[28] The European Union's General Data Protection Regulation seems to infer a "right to explanation." See Burt (2017).

aggregators and their related AI firms, with whom we have no relationship (for instance a contract), have more information about our preferences, concerns, and vulnerabilities than our priests, doctors, lawyers, or therapists. I propose that Governments should extend their existing off-line protections and standards against manipulation to also cover these data controllers which presently have the knowledge and proximity of a very intimate relationship without the governance and trust inherent to such relationships in the off-line market. I also propose several steps to protect citizens' autonomy and decrease user deception.

Regulating manipulation to protect consumer choice is not novel. What is unique now is that the current incarnation of manipulation online divorces the intimate knowledge of the target and power used to manipulate from a specific, ethically-regulated relationship as we usually find offline. Online we now have a situation where firms, with whom we have no relationship, have more information about our preferences, concerns, and vulnerabilities than our priests, doctors, lawyers, or therapists. In addition, these firms, such as ad networks, data brokers, and data aggregators, have an ability to reach specific targets due to the hypertargeting mechanisms available online. Yet, we are not privy to who has access to that information when businesses approach us with targeted product suggestions or advertising. These data brokers have the knowledge and proximity of an intimate relationship covering very personal parts of our lives without the governance and trust inherent to such relationships in the market. They clearly fail the transparency, stewardship, non-discrimination, autonomy, and fairness provisions of the G20 Principles.

Current Approach to Regulating Manipulation Online.

In the offline world sharing information with a particular market actor, such as a firm or individual, requires trust and other safeguards such as regulation, professional duties, contracts, negotiated alliances, nondisclosure agreements, etc. The point of such instruments is to share information within a (now legally binding) safe environment where the interests of the two actors are forced to be aligned. However, three facets of manipulation by data traffickers[29] – those in a position to covertly exploit the relative vulnerabilities or weaknesses of a person in order to usurp their decision making – strain our current mechanisms governing privacy and data. First, manipulation works by not being disclosed, thus making detection difficult and rendering the market ill-equipped to govern the behavior. Second, the type of manipulation described herein is performed by multiple economic actors including websites/apps, trackers, data aggregators, ad networks, and customer facing websites luring in the target. Third, data traffickers – who collect, aggregate, and sell consumer data – are the engine of manipulation of online consumers yet have no interaction, contract, agreement with individuals.

These three facets – manipulation is deceptive, shared between actors, and not visible by individuals – render the current mechanisms ineffective in governing the behavior or the actors. For example, GDPR is strained when attempting to limit a 'legitimate use' of data traffickers or data brokers who are looking to market products and services based on intimate knowledge. An individual has a right to the restriction of processing of information only when there are no legitimate grounds of the data controller. This makes GDPR fall short because legitimate interests can be broadly construed to include product placements and

[29] Lauren Scholz first used the term data traffickers, rather than data brokers, to describe firms that remain hidden yet traffic in such consumer data (Scholz 2019).

ads. And the manipulation of individuals has not been identified (yet) as diminishing a human right of freedom and autonomy. One fix is to more clearly link manipulation to individual autonomy, which would be seen as a human right that could trump even the legitimate interests of data traffickers.

A first step forward – Policy Goals

In general, the danger comes from using intimate knowledge about an individual and hyper-targeting to then manipulate them. The combination of individualized data and individualized targeting needs to be governed or limited:

1. **Protect Autonomy.** Manipulation is only possible because a market actor, here it is data brokers, has intimate knowledge of individuals as to what renders a target vulnerable in their decision making. The goal of governance would be to limit the use of intimate knowledge by making certain types of intimate knowledge either illegal or heavily governed. The combination of intimate knowledge with hyper- targeting of individuals should be more closely regulated than blanket targeting based on age and gender. Explicitly recognize individual autonomy, defined as the ability of individuals to be the authentic authors of their own decisions, as a legal right in order to protect individuals from manipulation done in the name of "legitimate interests" within the AI Principles.
2. **Expand Definitions of Intimate Knowledge.** One step would be to explicitly include inferences made about individuals as sensitive information within such existing regulations as GDPR (Wachter and Mittelstadt 2019). Sandra Wachter and Brent Mittelstadt have recently called on rights of access, notification, and correction for not only the data being collected but the possible inferences drawn from the data about individuals. These inferences would be considered intimate knowledge of individuals that could be used to manipulate them (e.g., whether someone is depressed or not based on their online activity). The inferences made by data traffickers based on a mosaic of information about individuals can constitute intimate knowledge as to who is vulnerable and when. Current regulatory approaches only include collected data as protected rather than the inferences drawn about individuals based on that data.
3. **Force Shared Responsibility.** Make customer-facing firms responsible for who they partner with to track users or to target users. Customer-facing websites and apps should be responsible for who is given access to their users' data – whether by sale or whether given access by placing trackers and beacons on their site. Third parties include all trackers, beacons, and third parties who purchase data or access to their users. Websites and apps would then be held responsible for whether they partner with firms that abide by GDPR standards, AI Principles, or new standards of care in the U.S. Holding customer facing firms responsible for how their partners (third party trackers) gather and use their users' data would be similar to holding a hospital responsible for how the patient is cared for by their contractors in the hospital or holding a car company responsible for a third party app in the car that then tracked your movements. This would force the customer-facing firm, with whom the individual has some influence, to make sure their users' interests are being respected.[30] The shift would be to have customer-facing firms be held responsible for how their partners (ad networks and media) treat their users.

[30] It is ironic that currently data traffickers can *sell* data to bad actors but they just can't have their data *stolen* by those same bad actors.

4. Expand the Definition of "Sold". Make sure all regulations include beacons and tracking companies in the any requirement to notify if user data is 'sold'.
5. Create a Fiduciary Duty for Data Brokers. there is a profound, yet relatively easy to implement, step to address this manipulation. G20 and other governments could make their AI Principles practical by extending the regulatory requirements they have on doctors, teachers, lawyers, government agencies and others who collect and act on the intimate data of individuals to also apply to data aggregators and their related AI implementations. Any actor who collects intimate data about an individual should be required to act on, share, or sell this data consistent with the interests of the person. This would force the alignment of interests between the target/consumer/user and the firm in the position to manipulate. Without any market pressures, data brokers who hold intimate knowledge of individuals, would need to be held to a fiduciary-like standard of care for how their data would be used.(Balkin 2015) This would mean data brokers would need to be responsible for how their products and services were used to possibly undermine the interests of the individuals.
6. Add Oversight. Add a GAAP-like governance structure over data traffickers and ad networks to ensure individualized data is not used to manipulate. With these economic actors well outside any market pressures, there are few pressures on the firms to align their actions with users' interests. A third step would be to make data traffickers abide by GAAP-like regulations. Recently McGeveran called for GAAP-like approach for data security, where companies would be held to a standard defined for all firm similar to the use of GAAP standards for accounting. However, the same concept should be applied to those who hold user data as to how they protect the data when profiting from it.[31] Audits could also be used in order to ensure data traffickers, who control and profit from intimate knowledge of individuals, are abiding by their standards. This would add a cost to those who traffic in customer vulnerabilities and provide a third party to verify that those holding intimate user data act in a way that is in the individuals' interests and protect firms from capitalizing on their vulnerabilities. A GAAP-line governance structure could be flexible enough to understand the market needs while still being responsive to protect individual rights and concerns.
7. Decrease Deception. Finally, manipulation works because the tactic is hidden from the target. The goal of governance would be to make the basis of manipulation open to the target and others. In other words, make the type of intimate knowledge used in targeting obvious and public. This could mean a notice (e.g., this ad was placed because the ad network believes you are diabetic) or this could mean a registry when hyper-targeting is used to allow others to analyze how and why individuals are being targeted. Registering would be particularly important for political advertising so that researchers and regulators can identify the basis for hyper-targeting. It should not be sufficient for an AI/data aggregator just to say "I am collecting all this information in the interests of the user to see tailored advertising." That is equivalent to a doctor saying "I collect all this data about a patient's health to ensure that the patient only knows the prescription I give the patient." Patients have to give permission for and are entitled to know what data is collected (indeed in many countries patients formally own their health data), what tests have been conducted and their results, what the diagnosis is – and they are entitled to a second opinion on the data. Similar sorts of transparency and accountability offline should apply online. In other areas, where a lawyer or

[31] McGeveran calls for a GAAP like approach for data security. Here we would have the same idea for data protection. Where standards are set and others must be certified to abide by them (McGeveran 2018).

realtor or financial advisor, has intimate knowledge and a conflict of interest (where they could profit in a way that is detrimental to their client), they must disclose their conflict and the basis for their conflict.

Putting a legal requirement for companies to use data in the interests of the data subject also demands an objective test to ensure that the interpretation of the "interests of the data subject" is not open to differing interpretations. Various entities and companies could claim to be acting in the individual's interest, as they define it, even if the individual believes they are not. We propose that the test be grounded in two existing bodies of law: conventions on human rights and law governing relationships between professionals and their data subjects (doctor-patient, lawyer-client etc.), particularly the law related to use of patient/client data so as not to manipulate or exploit the data subject.[32]

The same principle holds for data that is generated by material objects owned by the data subject. The IOT digital service provider, when different from the owner of the material objects, are to manage the IOT data flow in the interests of the data subject and the data subject needs to be given automatic access to the data generated by the relevant material objects. This data, along with associated terms and conditions, must be transparent and clear.

In the offline world, we have stressed the importance of clear relationships between people and those who have intimate information asymmetries over them. And we have developed safeguards to ensure that those gaining positions of power do not exploit vulnerabilities and weaknesses of individuals. The issues posed by vast data collection and hyper-targeted marketing and/or service delivery are a product of the global expanse of the Internet, social media and AI platforms. Furthermore, the ability of 'data traffickers' and their AI partners to leverage knowledge they have on almost every person on the Internet makes the scale of the public policy and political challenge worthy of Ministers and Heads of Government. As the growing "tech backlash" shows, there is political mobilization among citizens across the world for change. The innovation of this "apply the offline world rules to the online players" approach is that it does not require governments to educate or force citizens to change behaviors or desires. It puts the ethical and regulatory onus on the firms involved and holds them accountable.

Multilateral Governmental Responses to Date

The questions of the correct governance for Artificial Intelligence and its underlying Big Data have been discussed at national and dispersed international fora for several years. These include efforts by the Council

32 Some examination of this law can be found at https://ec.europa.eu/health/sites/health/files/cross_border_care/docs/2018_mapping_patientsrights_frep_en.pdf

of Europe[33], the Innovation Ministers of the G7[34] , the European Parliament [35] and the OECD.[36] In June 2019, China's Ministry of Science and Technology published on its website the Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence.[37] The same month, the G20 Trade Ministers and Digital Economy Ministers adopted a set of AI Principles[38] which drew from the OECD's principals and discussion of proposals from G20 engagement groups[39] These principles point to a more human-focused and ethical approach to guiding AI – but they are by necessity broad in tone and lacking in regulatory specifics. The G20 principles are attached as an Appendix A.

On 11 June 2020, United Nations Secretary General Guterres presented his *Roadmap on Digital Cooperation*.[40] One of the recommended actions is "Supporting global cooperation on artificial intelligence that is trustworthy, human-rights based, safe and sustainable and promotes peace."

In April 2021, the European Commission released its Proposal for consideration of the European Parliament and Council for the promotion and regulation of AI in Europe.[41] The media brief outlining the full package is attached as Appendix B

The proposal is more detailed than previous statements of principles. Among a range of issues, the draft regulations seek to cover facial recognition, autonomous driving, the use of AI in online advertising, automated hiring, and credit scoring. They seek to prohibit (at least in some ways) "high risk" applications of AI, including law enforcement real time use of AI for facial recognition in public spaces (but not its post-facto uses in a number of circumstances).

---

[33] https://rm.coe.int/algorithms-and-human-rights-study-on-the-human-rights-dimension-of- aut/1680796d10

[34] https://g7.gc.ca/en/g7-presidency/themes/preparing-jobs-future/g7-ministerial-meeting/chairs- summary/annex-b/

[35] Directorate-General for Parliamentary Research Services (European Parliament), A governance framework for algorithmic accountability and transparency see at https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf

[36] https://www.oecd.org/going-digital/ai/principles/

[37] See https://perma.cc/V9FL-H6J7

[38] See Annex to G20 Ministerial Statement on Trade and Digital Economy at https://www.mofa.go.jp/files/000486596.pdf

[39] For instance, see Paul Twomey. "Building on the Hamburg Statement and the G20 Roadmap for Digitalization: Toward a G20 framework for artificial intelligence in the workplace." At https://www.g20-insights.org/policy_briefs/building-on-the-hamburg-statement-and-the-g20-roadmap-for-digitalization-towards-a-g20-framework-for-artificial-intelligence-in-the-workplace/

[40] See https://www.un.org/en/content/digital-cooperation-roadmap/

[41] 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, COM/2021/206 final. See https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206

The proposed EU rules would also prohibit "AI-based social scoring for general purposes done by public authorities," as well as AI systems that target "specific vulnerable groups" in ways that would "materially distort their behavior" to cause "psychological or physical harm." This could stop the use of AI for credit scoring, hiring, or some forms of surveillance advertising.

Like the European Commission's proposal, the Feasibility Study of CAHAI lays forth quite detailed provisions to address specific rights in European law. CAHAI is we are working on elaboration of elements of binding and non-binding instruments, and according to the timeline set by the Committee of Ministers (governing body composed by Foreign Ministers of 47 member countries) negotiations on a treaty are to start before May 2022.

In July 2021, representatives of Member States on UNESCO agreed on a draft recommendation on AI governance, to be submitted to the General Conference of UNESCO Member States in November 2021 for adoption[42]. The objectives of the draft Recommendation are:

(a) to provide a universal framework of values, principles and actions to guide States in the formulation of their legislation, policies or other instruments regarding AI;

(b) to guide the actions of individuals, groups, communities, institutions and private sector companies to ensure the embedding of ethics in all stages of the AI system life cycle;

(c) to promote respect for human dignity and gender equality, to safeguard the interests of present and future generations, and to protect human rights, fundamental freedoms, and the environment and ecosystems in all stages of the AI system life cycle;

(d) to foster multi-stakeholder, multidisciplinary and pluralistic dialogue about ethical issues relating to AI systems; and

(e) to promote equitable access to developments and knowledge in the field of AI and the sharing of benefits, with particular attention to the needs and contributions of LMICs, including LDCs, LLDCs and SIDS.

The draft recommendations include a call for an international regulatory framework to ensure that AI benefits humanity as a whole and respect, protection and promotion of human dignity, human rights and fundamental freedoms.

Questions considered in discussion among the Sub Committee

Q What can be achieved in an international agreement? How can we achieve consensus of policy across the three dominant models of data governance: Enterprise-centred Internet (US), State-centred Internet (China) and Citizen-centred Internet (the EU and other OECD partner countries)?

[42] See https://en.unesco.org/artificial-intelligence/ethics

Q What level of further detail above the G20 principles could be adopted by a broad range of states?

Q What does an international agreement accept or ignore in the EU draft legislation?

Q How does this effort relate to the meetings of governmental experts and officials in Geneva under the auspices of the Convention on Certain Conventional Weapons[43] to continue trying to find consensus on next steps in regulating the next class of automated weapons?

Conclusions of the Sub Committee's video conference discussions

The Subcommittee concluded that the ambition of the Policy Lab should be an international accord with the greatest appeal to all countries (at least members of the UN). In this sense it should be more like the Kyoto Agreement or existing UN human rights treaties rather than a treaty with a more limited likely membership such as the treaty on the prohibition of nuclear weapons.

As we examined the various statements and draft multilateral documents on AI we realized that we were not comparing apples with apples, but rather we were dealing with more vague apples and quite specific pears. The challenge for our first Sub-Committee discussion was what mix of the two do we think is realistic.

To that end the Subcommittee recommended that the definition of human rights being protected by the proposed international accord be as widely accepted as possible. We focused on those outlined by the 1948 Universal Declaration of Human Rights. These could be expanded to include rights outlined in the nine core UN human rights treaties:

- the International Covenant on Civil and Political Rights (ICCPR)
- the International Covenant on Economic, Social and Cultural Rights (ICESCR)
- the International Convention on the Elimination of All Forms of Racial Discrimination (CERD)
- the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW)
- the Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (CAT)
- International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families (ICMW)
- the Convention on the Rights of the Child (CRC)
- the Convention on the Rights of Persons with Disabilities (CRPD).[44]

The Subcommittee discussion recognised that even these expanded treaties are controversial in some quarters and can be see as being deriving from the basic 30 principles outlined in the Universal Declaration. Hence our focus on the 1948 document. Indeed we concluded that 28 of the rights outlined could be negatively affected by AI, while 26 could be promoted through careful application of AI

[43] See https://www.un.org/disarmament/
[44] See https://www.ohchr.org/en/professionalinterest/pages/coreinstruments.aspx

applications. (See Appendix C). Further, we though that specific focus should also be made on limiting the use of AI to manipulate users, particularly the vulnerable.

We also considered that the move to an international accord should not be held hostage to the difficulty governments are having to find consensus on regulating the next class of autonomous weapons. To the degree it is possible the two discussions should not be linked.

As for what wording to move forward on for the negotiation of an international accord, the subcommittee discussion considered that the G20 statement should be considered a good starting position (perhaps augmented by some of the wording of the UNESCO recommendation if and how it is approved by the General Assembly). The members in discussion suspected that the EU draft legislation and even the CAHAI documents may fail to attract the universal approval we think is necessary for a global accord.

Finally, the discussion in the Sub Committee suggested that an international accord should also call for transparency and a call for pause and international review (if not a total moratorium) on the transition to General Intelligence by AI initiatives in their jurisdictions. An unrestricted move to General Intelligence would in our view pose a very significant potential threat to human rights.

Appendix A

G20 AI Principles

The G20 supports the Principles for responsible stewardship of Trustworthy AI in Section 1 and takes note of the Recommendations in Section 2.

Section 1: Principles for responsible stewardship of trustworthy AI

1.1. Inclusive growth, sustainable development and well-being
Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as augmenting human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being.

1.2. Human-centered values and fairness
a) AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognized labor rights.
b) To this end, AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of art.

1.3. Transparency and explainability
AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art:
i. to foster a general understanding of AI systems;
ii. to make stakeholders aware of their interactions with AI systems, including in the workplace;
iii. to enable those affected by an AI system to understand the outcome; and,
iv. to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.

1.4. Robustness, security and safety
a) AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk.
b) To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art.
c) AI actors should, based on their roles, the context, and their ability to act, apply a systematic

risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.

1.5. Accountability
AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art.

Section 2: National policies and international co-operation for trustworthy AI

2.1. Investing in AI research and development
a) Governments should consider long-term public investment, and encourage private investment, in research and development, including inter-disciplinary efforts, to spur innovation in trustworthy AI that focus on challenging technical issues and on AI-related social, legal and ethical implications and policy issues.
b) Governments should also consider public investment and encourage private investment in open datasets that are representative and respect privacy and data protection to support an environment for AI research and development that is free of inappropriate bias and to improve interoperability and use of standards.

2.2. Fostering a digital ecosystem for AI
Governments should foster the development of, and access to, a digital ecosystem for trustworthy AI. Such an ecosystem includes in particular digital technologies and infrastructure, and mechanisms for sharing AI knowledge, as appropriate. In this regard, governments should consider promoting mechanisms, such as data trusts, to support the safe, fair, legal and ethical sharing of data.

2.3 Shaping an enabling policy environment for AI
a) Governments should promote a policy environment that supports an agile transition from the research and development stage to the deployment and operation stage for trustworthy AI systems. To this effect, they should consider using experimentation to provide a controlled environment in which AI systems can be tested, and scaled-up, as appropriate.
b) Governments should review and adapt, as appropriate, their policy and regulatory frameworks and assessment mechanisms as they apply to AI systems to encourage innovation and competition for trustworthy AI.

2.4. Building human capacity and preparing for labor market transformation
a) Governments should work closely with stakeholders to prepare for the transformation of the world of work and of society. They should empower people to effectively use and interact with AI systems across the breadth of applications, including by equipping them with the necessary skills.
b) Governments should take steps, including through social dialogue, to ensure a fair transition for workers as AI is deployed, such as through training programs along the working life, support for

those affected by displacement, and access to new opportunities in the labor market.

c) Governments should also work closely with stakeholders to promote the responsible use of AI at work, to enhance the safety of workers and the quality of jobs, to foster entrepreneurship and productivity, and aim to ensure that the benefits from AI are broadly and fairly shared.

2.5. International co-operation for trustworthy AI
a) Governments, including developing countries and with stakeholders, should actively cooperate to advance these principles and to progress on responsible stewardship of trustworthy AI.
b) Governments should work together in the OECD and other global and regional fora to foster the sharing of AI knowledge, as appropriate. They should encourage international, cross-sectoral and open multi-stakeholder initiatives to garner long-term expertise on AI.
c) Governments should promote the development of multi-stakeholder, consensus-driven global technical standards for interoperable and trustworthy AI.
d) Governments should also encourage the development, and their own use, of internationally comparable metrics to measure AI research, development and deployment, and gather the evidence base to assess progress in the implementation of these principles.

Appendix B

Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence

Brussels, 21 April 2021

The Commission proposes today new rules and actions aiming to turn Europe into the global hub for trustworthy Artificial Intelligence (AI). The combination of the first-ever legal framework on AI and a new Coordinated Plan with Member States will guarantee the safety and fundamental rights of people and businesses, while strengthening AI uptake, investment and innovation across the EU. New rules on Machinery will complement this approach by adapting safety rules to increase users' trust in the new, versatile generation of products.

Margrethe **Vestager**, Executive Vice-President for a Europe fit for the Digital Age, said: *"On Artificial Intelligence, trust is a must, not a nice to have. With these landmark rules, the EU is spearheading the development of new global norms to make sure AI can be trusted. By setting the standards, we can pave the way to ethical technology worldwide and ensure that the EU remains competitive along the way. Future-proof and innovation-friendly, our rules will intervene where strictly needed: when the safety and fundamental rights of EU citizens are at stake."*

Commissioner for Internal Market Thierry **Breton** said: *"AI is a means, not an end. It has been around for decades but has reached new capacities fueled by computing power. This offers immense potential in areas as diverse as health, transport, energy, agriculture, tourism or cybersecurity. It also presents a number of risks. Today's proposals aim to strengthen Europe's position as a global hub of excellence in AI from the lab to the market, ensure that AI in Europe respects our values and rules, and harness the potential of AI for industrial use."*

The new **AI regulation** will make sure that Europeans can trust what AI has to offer. Proportionate and flexible rules will address the specific risks posed by AI systems and set the highest standard worldwide. The **Coordinated Plan** outlines the necessary policy changes and investment at Member States level to strengthen Europe's leading position in the development of human-centric, sustainable, secure, inclusive and trustworthy AI.

The European approach to trustworthy AI

The new rules will be applied directly in the same way across all Member States based on a future- proof definition of AI. They follow a risk-based approach:

Unacceptable risk: AI systems considered a clear threat to the safety, livelihoods and rights of people **will be banned**. This includes AI systems or applications that manipulate human behaviour to circumvent users' free will (e.g. toys using voice assistance encouraging dangerous behaviour of minors) and systems that allow 'social scoring' by governments.

High-risk: AI systems identified as high-risk include AI technology used in:

- **Critical infrastructures** (e.g. transport), that could put the life and health of citizens at risk;
- **Educational or vocational training**, that may determine the access to education and professional course of someone's life (e.g. scoring of exams);
- **Safety components of products** (e.g. AI application in robot-assisted surgery);
- **Employment, workers management and access to self-employment** (e.g. CV-sorting software for recruitment procedures);

- **Essential private and public services** (e.g. credit scoring denying citizens opportunity to obtain a loan);
- **Law enforcement** that may interfere with people's fundamental rights (e.g. evaluation of the reliability of evidence);
- **Migration, asylum and border control management** (e.g. verification of authenticity of travel documents);
- **Administration of justice** and **democratic processes** (e.g. applying the law to a concrete
- set of facts).
- High-risk AI systems will be subject to **strict obligations** before they can be put on the market:
- **Adequate risk assessment and mitigation systems;**
- **High quality of the datasets** feeding the system to minimise risks and discriminatory outcomes;
- **Logging of activity to ensure traceability of results**;
- **Detailed documentation** providing all information necessary on the system and its purpose for authorities to assess its compliance;
- **Clear and adequate information** to the user;
  Appropriate human oversight measures to minimise risk;
  High level of **robustness**, **security** and **accuracy**.

In particular, **all remote biometric identification** systems are considered high risk and subject to strict requirements. Their live use in publicly accessible spaces for law enforcement purposes is prohibited in principle. Narrow exceptions are strictly defined and regulated (such as where strictly necessary to search for a missing child, to prevent a specific and imminent terrorist threat or to detect, locate, identify or prosecute a perpetrator or suspect of a serious criminal offence). Such use is subject to authorisation by a judicial or other independent body and to appropriate limits in time, geographic reach and the data bases searched.

Limited risk, i.e. AI systems with specific transparency obligations: When using AI systems such as chatbots, users should be aware that they are interacting with a machine so they can take an informed decision to continue or step back.

Minimal risk: The legal proposal allows the free use of applications such as AI-enabled video games or spam filters. The vast majority of AI systems fall into this category. The draft Regulation does not intervene here, as these AI systems represent only minimal or no risk for citizens' rights or safety.

In terms of governance, the Commission proposes that national competent market surveillance authorities supervise the new rules, while the creation of a **European Artificial Intelligence Board** will facilitate their implementation, as well as drive the development of standards for AI. Additionally, voluntary codes of conduct are proposed for non-high-risk AI, as well as regulatory sandboxes to facilitate responsible innovation.

The European approach to excellence in AI

Coordination will strengthen Europe's leading position in human-centric, sustainable, secure, inclusive and trustworthy AI. To remain globally competitive, the Commission is committed to fostering innovation in AI technology development and use across all industries, in all Member States.

First published in 2018 to define actions and funding instruments for the development and uptake of AI, the **Coordinated Plan on AI** enabled a vibrant landscape of national strategies and EU funding for public-private partnerships

and research and innovation networks. The comprehensive update of the Coordinated Plan proposes concrete joint actions for collaboration to ensure all efforts are aligned with the European Strategy on AI and the European Green Deal, while taking into account new challenges brought by the coronavirus pandemic. It puts forward a vision to accelerate investments in AI, which can benefit the recovery. It also aims to spur the implementation of national AI strategies, remove fragmentation, and address global challenges.

The updated Coordinated Plan will use funding allocated through the **Digital Europe** and **Horizon Europe** programmes, as well as the **Recovery and Resilience Facility** that foresees a 20% digital expenditure target, and **Cohesion Policy** programmes, to:

- **Create enabling conditions for AI's development** and uptake through the exchange of policy insights, data sharing and investment in critical computing capacities;
- **Foster AI excellence** 'from the lab to the market' by setting up a public-private partnership, building and mobilising research, development and innovation capacities, and making testing and experimentation facilities as well as digital innovation hubs available to SMEs and public administrations;
- **Ensure that AI works for people** and is a force for good in society by being at the forefront of the development and deployment of trustworthy AI, nurturing talents and skills by supporting traineeships, doctoral networks and postdoctoral fellowships in digital areas, integrating Trust into AI policies and promoting the European vision of sustainable and

  trustworthy AI globally;
- **Build strategic leadership** in high-impact sectors and technologies including environment by focusing on AI's contribution to sustainable production, health by expanding the cross-border exchange of information, as well as the public sector, mobility, home affairs and agriculture, and Robotics.

The European approach to new machinery products

Machinery products cover an extensive range of consumer and professional products, from robots to lawnmowers, 3D printers, construction machines, industrial production lines. The Machinery Directive, replaced by the new Machinery Regulation, defined health and safety requirements for machinery. This new Machinery Regulation will ensure that the new generation of machinery guarantees the safety of users and consumers, and encourages innovation. While the AI Regulation will address the safety risks of AI systems, the new Machinery Regulation will ensure the safe integration of the AI system into the overall machinery. Businesses will need to perform only one single conformity assessment.

Additionally, the new Machinery Regulation will respond to the market needs by bringing greater legal clarity to the current provisions, simplifying the administrative burden and costs for companies by allowing digital formats for documentation and adapting conformity assessment fees for SMEs, while ensuring coherence with the EU legislative framework for products.

Next steps

The European Parliament and the Member States will need to adopt the Commission's proposals on a European approach for Artificial Intelligence and on Machinery Products in the ordinary legislative procedure. Once adopted, the Regulations will be directly applicable across the EU. In parallel, the Commission will continue to collaborate with Member States to implement the actions announced in the Coordinated Plan.

Background

For years, the Commission has been facilitating and enhancing cooperation on AI across the EU to boost its competitiveness and ensure trust based on EU values.

Following the publication of the European Strategy on AI in 2018 and after extensive stakeholder consultation, the High-Level Expert Group on Artificial Intelligence (HLEG) developed Guidelines for Trustworthy AI in 2019, and an Assessment List for Trustworthy AI in 2020. In parallel, the first Coordinated Plan on AI was published in December 2018 as a joint commitment with Member States.

The Commission's White Paper on AI, published in 2020, set out a clear vision for AI in Europe: an ecosystem of excellence and trust, setting the scene for today's proposal. The public consultation on the White Paper on AI elicited widespread participation from across the world. The White Paper was accompanied by a 'Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics' concluding that the current product safety legislation contains a number of gaps that needed to be addressed, notably in the Machinery Directive.

For More Information

New rules for Artificial Intelligence – Questions and Answers New rules for Artificial Intelligence – Facts page

Communication on Fostering a European approach to Artificial Intelligence Regulation on a European approach for Artificial Intelligence

New Coordinated Plan on Artificial Intelligence Regulation on Machinery Products

EU-funded AI projects

Appendix C

List of 30 basic human rights and their possible impact by AI

The Universal Declaration of Human Rights was approved by the United Nations General Assembly at the Palais de Chaillot in Paris, France on 10 December 1948. Of the then 58 members of the United Nations, 48 voted in favor, none against, eight abstained, and two did not vote.

This declaration consists of 30 articles affirming an individual's rights.

Beside each right I have placed my judgement as to how a right could be affected by AI applications. You will see that the majority could be both negatively impacted and also positively impacted – a not unusual consequence of such a powerful enabling technology.

Paul Twomey
1 September 2021

1. All human beings are free and equal Negatively

All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.

2. No discrimination Both Negatively or Positively

Everyone is entitled to all the rights and freedoms, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status. Furthermore, no distinction shall be made on the basis of the political, jurisdictional or international status of the country or territory to which a person belongs.

3. Right to life Both Negatively or Positively

Everyone has the right to life, liberty and security of person.

4. No slavery Both Negatively or Positively

No one shall be held in slavery or servitude; slavery and the slave trade shall be prohibited in all their forms.

5. No torture and inhuman treatment Both Negatively or Positively

No one shall be subjected to torture or to cruel, inhuman or degrading treatment or punishment.

6. Same right to use law Both Negatively or Positively

Everyone has the right to recognition everywhere as a person before the law.

7. Equal before the law Both Negatively or Positively

All are equal before the law and are entitled without any discrimination to equal protection of the law. All are entitled to equal protection against any discrimination in violation and against any incitement to such discrimination.

8. Right to treated fair by court Both Negatively or Positively

Everyone has the right to an effective remedy by the competent national tribunals for acts violating the fundamental rights granted him by the constitution or by law.

9. No unfair detainment Both Negatively or Positively

No one shall be subjected to arbitrary arrest, detention or exile.

10. Right to trial Both Negatively or Positively

Everyone is entitled in full equality to a fair and public hearing by an independent and impartial tribunal, in the determination of his rights and obligations and of any criminal charge against him.

11. Innocent until proved guilty Negatively

Everyone charged with a penal offence has the right to be presumed innocent until proved guilty according to law in a public trial at which he has had all the guarantees necessary for his defence. No one shall be held guilty of any penal offence on account of any act or omission which did not constitute a penal offence, under national or international law, at the time when it was committed.

12. Right to privacy Both Negatively or Positively

No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.

13. Freedom to movement and residence Both Negatively or Positively

Everyone has the right to freedom of movement and residence within the borders of each state. Everyone has the right to leave any country, including his own, and to return to his country.

14. Right to asylum Both Negatively or Positively

Everyone has the right to seek and to enjoy in other countries asylum from persecution. This right may not be invoked in the case of prosecutions genuinely arising from non-political crimes or from acts contrary to the purposes and principles of the United Nations.

15. Right to nationality Both Negatively or Positively

Everyone has the right to a nationality. No one shall be arbitrarily deprived of his nationality nor denied the right to change his nationality

16. Rights to marry and have family Both Negatively or Positively

Men and women of full age, without any limitation due to race, nationality or religion, have the right to marry and to found a family. They are entitled to equal rights as to marriage, during marriage and at its dissolution. Marriage shall be entered into only with the free and full consent of the intending spouses. The family is the natural and fundamental group unit of society and is entitled to protection by society and the State.

17. Right to own things Both Negatively or Positively

Everyone has the right to own property alone as well as in association with others. No one shall be arbitrarily deprived of his property.

18. Freedom of thought and religion Both Negatively or Positively

Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief, and freedom, either alone or in community with others and in public or private, to manifest his religion or belief in teaching, practice, worship and observance.

19. Freedom of opinion and expression Both Negatively or Positively

Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.

20. Right to assemble Both Negatively or Positively

Everyone has the right to freedom of peaceful assembly and association. No one may be compelled to belong to an association.

21. Right to democracy Both Negatively or Positively

Everyone has the right to take part in the government of his country, directly or through freely chosen representatives. Everyone has the right of equal access to public service in his country.

22. Right to social security Both Negatively or Positively

Everyone, as a member of society, has the right to social security and is entitled to realization, through national effort and international co-operation and in accordance with the organization and resources of each State, of the economic, social and cultural rights indispensable for his dignity and the free development of his personality.

23. Right to work Both Negatively or Positively

Everyone has the right to work, to free choice of employment, to just and favourable conditions of work and to protection against unemployment. Everyone, without any discrimination, has the right to equal pay for equal work. Everyone has the right to form and to join trade unions for the protection of his interests.

24. Right to rest and holiday Both Negatively or Positively

Everyone has the right to rest and leisure, including reasonable limitation of working hours and periodic holidays with pay.

25. Right of social service Both Negatively or Positively

Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of livelihood in circumstances beyond his control. Motherhood and childhood are entitled to special care and assistance. All children shall enjoy the same social protection.

26. Right to education Both Negatively or Positively

Everyone has the right to education. Education shall be free, at least in the elementary and fundamental stages. Elementary education shall be compulsory. Technical and professional education shall be made generally available and higher education shall be equally accessible to all on the basis of merit.

27. Right of cultural and art Both Negatively or Positively

Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits. Everyone has the right to the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author.

28. Freedom around the world Both Negatively or Positively

Everyone is entitled to a social and international order in which the rights and freedoms set forth in this Declaration can be fully realized.

29. Subject to law Both Negatively or Positively

Everyone has duties to the community in which alone the free and full development of his personality is possible. In the exercise of his rights and freedoms, everyone shall be subject only to such limitations as are determined by law solely for the purpose of securing due recognition and respect for the rights and freedoms of others and of meeting the just requirements of morality, public order and the general welfare in a democratic society.

30. Human rights can't be taken away Negatively

Nothing in this Declaration may be interpreted as implying for any State, group or person any right to engage in any activity or to perform any act aimed at the destruction of any of the rights and freedoms set forth herein.

Works Cited

Acquisti, Allesandro, and Christina M. Fong. 2015. "An Experiment in Hiring Discrimination via Online Social Networks." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2031979.

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

Balkin, Jack M. 2015. "Information Fiduciaries and the First Amendment." *UCDL Rev.* 49: 1183.

Barghava, Rahul. 2017. "The Algorithms Aren't Biased, We Are." *MIT Media Lab* (blog), January 3. https://medium.com/mit-media-lab/the-algorithms-arent-biased-we-are-a691f5f6f6f2.

Barocas, Solon, and Andrew D. Selbst. 2016. "Big Data's Disparate Impact." *California Law Review* 104: 671–732.

British Columbia First Nations Data Governance Initiative. 2017. *Decolonizing Data: Indigenous Data Sovereignty Primer*. April.

Burt, Andrew. 2017. "Is there a 'right to explanation' for machine learning in the GDPR?" IAPP *Privacy Tech* (blog), June 2. International Association of Privacy Professionals. https://iapp.org/news/a/is-there-a-right-to-explanation-for-machine-learning-in-the-gdpr/.

Cellan-Jones, Rory. 2014. "Stephen Hawking warns artificial intelligence could end mankind." BBC News, December 2.. www.bbc.com/news/technology-30290540.

Clifford, Catherine . 2017. "Hundreds of A.I. experts echo Elon Mush, Stephen Hawking in call for a ban on killer robots." CNBC, November 8. https://www.cnbc.com/2017/11/08/ai-experts-join-elon-musk-stephen-hawking-call-for-killer-robot-ban.html.

Englehardt, Steven, and Arvind Narayanan. 2016. "Online Tracking: A 1-Million-Site Measurement and Analysis." http://randomwalker.info/publications/OpenWPM_1_million_site_tracking_measurement.pdf.

Eubanks, Virgina. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* New York, NY: St. Martin's Press.

Executive Office of the President. 2014. *Big Data: Seizing Opportunities, Preserving Values*. Washington, DC: The White House. https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.

Frey, Carl Benedikt and Michael A. Osborne. 2013. *The Future of Employment: How Susceptible Are Jobs to Computerisation?* September 17. Oxford, UK: Oxford Martin Programme on Technology and Employment. www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf.

G20. 2017a. *G20 Digital Economy Ministerial Conference. Düsseldorf 6-7 April 2017.* Declaration of the Ministers Responsible for the Digital Economy. Federal Ministry for Economic Affairs and Energy. www.bmwi.de/Redaktion/DE/Downloads/G/g20-digital-economy-ministerial-declaration-english-version.pdf?__blob=publicationFile&v=12.

G20. 2017b. "A ROADMAP for Digitalisation: Policies for a Digital Future." Annex paper 1 to the Declaration of the Ministers responsible for the Digital Economy. In *G20 Digital Economy Ministerial Conference. Düsseldorf 6-7 April 2017*, 10–15. www.bmwi.de/Redaktion/DE/Downloads/G/g20-digital-economy-ministerial-declaration-english-version.pdf?__blob=publicationFile&v=12.

G20. 2017c. "G20 Leaders' Declaration: Shaping an interconnected world." G20 Germany 2017 meetings, Hamburg, July 7-8. www.g20germany.de/Content/EN/_Anlagen/G20/G20-leaders-declaration.pdf;jsessionid=0C08AA235271BF43ECBB08BA059EE5B7.s6t2?__blob=publicationFile&v=11.

Gangadharan, Seeta P., Virginia Eubanks and Solon Barocas, eds. 2014. *Data and Discrimination: Collected Essays*. Washington, DC: New America. www.newamerica.org/oti/policy-papers/data-and-discrimination/.

Global Commission on Internet Governance. 2016. *One Internet: Final Report of the Global Commission on Internet Governance.* Waterloo, ON: CIGI.. www.cigionline.org/publications/one-internet.

Hurley, Dan. 2018. "Can an Algorithm Tell When Kids Are in Danger?" *New York Times*, January 2. www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html.

Jaume-Palasí, Lorena and Matthias Spielkamp. 2017. "Ethics and algorithmic processes for decision making and decision support." AlgorithmWatch Working Paper No. 2, June 1. Berlin, Germany: AlgorithmWatch. https://algorithmwatch.org/en/ethics-and-algorithmic-processes-for-decision-making-and-decision-support/.

Kirchner, Lauren. 2017. "New York City moves to create accountability for algorithms." *Ars Technica*, December 19. https://arstechnica.com/tech-policy/2017/12/new-york-city-moves-to-create-accountability-for-algorithms/.

KPMG International. 2016. "Rise of the humans: The integration of digital and human labor." KPMG International Cooperative, November. https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2016/11/rise-of-the-humans.pdf.

Kroll, Joshua A., Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson and Harlan Yu. 2017. "Accountable algorithms." *University of Pennsylvania Law Review* 165: 633–705.

Lohr, Steve. 2017. "A.I. will transform the Economy. But How Much, and How Soon?" *New York Times*, November 30. www.nytimes.com/2017/11/30/technology/ai-will-transform-the-economy-but-how-much-and-how-soon.html.

Madden, Mary, Michele Gilman, Karen Levy and Alice Marwick. 2017. "Privacy, Poverty, and Big Data: A Matrix of Vulnerabilities for Poor Americans." *Washington University Law Review* 95 (1): 53–125.

Manyika, James, Susan Lund, Michael Chui, Jacques Bughin, Jonathan Woetzel, Parul Batra, Ryan Ko and Saurabh Sanghvi. 2017. *Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation*. San Francisco, CA: McKinsey Global Institute. www.mckinsey.com/featured-insights/future-of-organizations-and-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages.

McGeveran, William. 2018. "The Duty of Data Security." *Minn. L. Rev.* 103: 1135.
Scholz, Lauren Henry. 2019. "Privacy Remedies." *Indiana Law Journal*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3159746

Microsoft. 2018. *The Future Computed: Artificial Intelligence and its role in society*. Redmond, WA: Microsoft. https://blogs.microsoft.com/uploads/2018/02/The-Future-Computed_2.8.18.pdf.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, NY: New York University Press.

Obar, Jonathan A. and Anne Oeldorf-Hirsch. 2016. "The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2757465.

Obar, Jonathan, and Brenda McPhail. 2018. "Preventing Big Data Discrimination in Canada: Addressing Design, Consent and Sovereignty Challenges." In *Data Governance in the Digital Age*: *Special Report*, 56–64. Waterloo, ON: CIGI. https://www.cigionline.org/publications/data-governance-digital-age.

O'Neil, Cathy. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY: Broadway Books.

Ovenden, James. 2016. "AI In Developing Countries: Artificial intelligence isn't just for self driving cars." Innovation Enterprise, October 6. https://channels.theinnovationenterprise.com/articles/ai-in-developing-countries.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms that Control Money and Information.* Cambridge, MA: Harvard University Press.

Purdy, Mark, and Paul Daugherty. 2016. "Why Artificial Intelligence Is the Future of Growth." Accenture Institute for High Performance, September 28. www.accenture.com/us-en/insight-artificial-intelligence-future-growth.

Reidenberg, Joel R., Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T. Graves, Fei Liu, Aleecia McDonald, Thomas B. Norton, Rohan Ramanath, N. Cameron Russell, Norman Sadeh and Florian Schaub. 2015. "Disagreeable Privacy Policies: Mismatches Between Meaning and Users'

Understanding." *Berkeley Technology Law Journal* 30 (1): 39–68. https://scholarship.law.berkeley.edu/cgi/viewcontent.cgi?article=2053&context=btlj.

Sandvig, Christian, Kevin Hamilton, Karrie Karahalios and Cedric Langbort. 2016. "When the Algorithm Itself Is a Racist: Diagnosing Ethical Harm in the Basic Components of Software." *International Journal of Communication* 10: 4972–90. http://social.cs.uiuc.edu/papers/pdfs/Sandvig-IJoC.pdf.

Scannell, R. Joshua. 2016. "Broken Windows, Broken Code." Reallifemag.com, August 29. http://reallifemag.com/broken-windows-broken-code/.

Solove, Daniel J. 2013. "Introduction: Privacy Self-Management and the Consent Dilemma." *Harvard Law Review* 126: 1880–1903. https://pdfs.semanticscholar.org/809c/bef85855e4c5333af40740fe532ac4b496d2.pdf.

Susser, Daniel, Beate Roessler, and Helen Nissenbaum. 2019. "Technology, Autonomy, and Manipulation." *Internet Policy Review* 8 (2).

The Economist. 2018. "AI-Spy: The workplace of the future." March 28. https://www.economist.com/leaders/2018/03/28/the-workplace-of-the-future.

Turow, Joseph. 2011. *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. New Haven, CT: Yale University Press.

UNI Global Union. 2017. *Top 10 Principles for Ethical Artificial Intelligence*. Nyon, Switzerland: UNI Global Union. www.thefutureworldofwork.org/opinions/10-principles-for-ethical-ai/.UNI Global Union. 2018. *Top 10 Principles for Workers' Data Privacy and Protection*. Nyon, Switzerland. www.thefutureworldofwork.org/docs/10-principles-for-workers-data-rights-and-privacy/.

Wachter, Sandra, and Brent Mittelstadt. 2019. "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI." *Columbia Business Law Review*

Wexler, Rebecca. 2017. "When a Computer Program Keeps You in Jail." New York Times, June 13. www.nytimes.com/2017/06/13/opinion/how-computers-are-harming-criminal-justice.html.

[i] Trewin, S., Basson, S., Muller, M., Branham, S., Treviranus, J., Gruen, D., Hebert, D., Lyckowski, N. and Manser, E., 2019. Considerations for AI fairness for people with disabilities. AI Matters, 5(3), pp.40-63.