

Paper Disclaimer

Draft – For discussion purposes only; does not represent the views of the G7 or its members. This discussion paper was drafted to guide the discussion during the breakout sessions at the December 6th, 2018, G7 Multistakeholder Conference on Artificial Intelligence in Montreal, Canada.

Discussion Paper for Breakout Session

Theme 1: AI for Society

Inclusion in AI Development and Deployment
Creating applications that address key socio-economic challenges
and serve all members of society positively

G7 Multistakeholder Conference on Artificial Intelligence
December 6, 2018, Montreal

Executive Summary

Artificial intelligence (AI) opens up wonderful opportunities, but democratic societies will need to address major challenges to ensure that the positive effects of algorithm development and use are shared equitably. In particular, we must ensure that AI growth does not amplify existing prejudice, or lead to the exclusion of vulnerable populations.

This short paper emphasizes the benefits of inclusion and diversity. It is also key to try to raise the numeracy and digital literacy levels of *all* citizens, and to improve access to key resources (including venture capital, infrastructure, databases, training).

There may be disparities in how AI-generated productivity gains are distributed (e.g. only rich countries or hospitals may have access to the latest tools for early detection of disease). As such, initiatives will need to be undertaken to promote the delivery of applications that benefit excluded groups (e.g. by promoting broad strategic engagement across government, developing technical solutions to social and ethical problems).

Even well-designed AI tools might have negative effects on the long-term economic and social inclusion of the world's population, e.g. automation-led job displacement. Countries will therefore be led to invest in the development of new industries, devise reskilling schemes, consider a broader social safety net, and modernize national labour policies.

There are also risks that the development of AI technologies may inadvertently lead to the exclusion of some groups. Researchers and private sector players have started to develop tools to mitigate such risks, but more will be needed to safeguard populations (e.g. closer involvement of end users during the design and implementation phases, ensuring adequate governance mechanisms). Finally, it is important to bear in mind that AI can and indeed should be part of the solution by revealing and disarming biases in existing systems and processes.

Discussion Questions for the Breakout Session

AI is sometimes framed as 'a space race' for the 21st century, but the major difference is that AI has the potential to benefit the whole humanity, in health, transportation, security, and environment. If we must ensure an inclusive development of AI at home, we should also bring nations together to achieve such aims as quickly as possible. We suggest using the discussion time to explore the following questions:

- Q1: How do we make sure gender, social and cultural diversity fuel AI design and development?
- Q2: What are the best practice examples of inclusion and diversity in AI and how can we replicate them?
- Q3: What can government, academia, and industry do to promote an inclusive use of AI technology that will benefit a diverse society?

Introduction

What is AI?

“In the broadest sense, AI refers to machines that can learn, reason, and act for themselves. They can make their own decisions when faced with new situations, in the same way that humans and animals can. As it currently stands, the vast majority of the AI advancements and applications you hear about refer to a category of algorithms known as machine learning. These algorithms use statistics to find patterns in massive amounts of data. They then use those patterns to make predictions on things like what shows you might like on Netflix, what you’re saying when you speak to Alexa, or whether you have cancer based on your MRI¹.”

Why is inclusion important?

In our democratic societies, everyone deserves proper treatment and basic rights. AI presents many wonderful opportunities, but we must also address important concerns. Here we highlight:

1. **Shared burdens:** algorithms are prone to replicating and reinforcing inappropriate bias for several reasons: training on biased datasets, input from designers or developers who are biased, or simply through lack of careful consideration to prevent bias. We must make sure that this does not lead to burdens falling unfairly on particular subgroups of society.
2. **Shared benefits:** the economic growth resulting from AI development could lead to benefits for only a few individuals and companies, while disrupting jobs for many. We must ensure that benefits are shared equitably between citizens and businesses in all countries.
3. **Homogeneity:** the developers of our AI systems are predominantly white and male yet these systems affect all of society. We must strive to attract a strong, inclusive community of those who build AI systems.
4. **Governance:** algorithmic systems raise concerns about accountability, particularly when it comes to decision-making, which impacts our lives. We should ensure that humans maintain control with appropriate transparency.

What is inclusion in the context of AI?

Ensuring the benefits of AI are felt by everyone

The global AI market is worth approximately USD 7.35bn² and will grow quickly in the future. This means that AI workers will be in high demand and will be well paid. It also means that countries and regions will benefit directly, economically, from fostering the development of competitive firms in the AI field. However, at the moment not all groups, countries and regions are equally represented in the AI industry. For example:

Draft – For discussion purposes only; does not represent the views of the G7 or its members.

- Colorintech, a non-profit working to promote inclusion in the UK's digital economy, found that only 2.6% of leadership positions are held by ethnic minorities. However, it should be noted that “when it comes to employment at all levels in the tech business, the UK performs better than the rest of the industry. Of those in digital technology jobs, 15% are of black, Asian and ethnic minority backgrounds” according to a 2018 TechNation report.
- WIRED estimated the diversity of leading machine learning researchers, and found that only 12% were women.³
- Venture capitalist Kai-Fu Lee warns that “most of the money being made from artificial intelligence will go to the United States and China [because] AI is an industry in which strength begets strength: the more data you have, the better your product; the better your product, the more data you can collect.”⁴

Possible solutions to make the AI sector more inclusive include ensuring that industry, and developers in particular, are aware of the importance and need for diversity. Other solutions include increasing the levels of numeracy and digital literacy of respective young students, and making key resources (e.g. venture capital, infrastructure, databases, language corpora, GPUs) more readily available in developing countries or regions.

For example, the BBC's “micro: bits campaign” made 90% of 11-year-old pupils feel “anyone can code” and 39% of girls (a jump of 16 percentage points) say they “would definitely do ICT/Computer Science as a subject option in the future.”⁵ In the U.S., AI4ALL, a non-profit, works to increase diversity and inclusion in AI by supporting educative and mentorship programs that give underrepresented high school students early exposure to AI for social good⁶. France's AI strategy states that incentives like accreditations or grants could be used to reward universities and business schools that achieve greater female enrolment rates⁷. Canadian start-up Element AI launched a new program in London to facilitate the “connection of leading AI scientists and engineers in developing nations through fellowships and visiting scholar programs⁸.” And in a different vein, Italy's AI White Paper on AI states that the country will need resources in Italian language (such as digital lexicons or annotated corpora distributed with open licenses) to develop some of the AI technologies it will need for its economic development and to “avoid being excluded from the AI race.”

Such work should rest in great part on the use of ‘on-the-ground’ expertise that arises from citizen participation. Citizen engagement is particularly useful when it comes to solving some of the issues regarding the development of ethical, safe and inclusive AI systems. Citizen centric groups can achieve this level of epistemic diversity, which often is very hard to engineer in high-level expert groups. An example of this can be found in the Montreal Declaration for a responsible development of AI, which relied on a participatory process of ‘co-construction.’ Over eight months, 15 workshops involving over 500 citizens, experts, and stakeholders have been organized in Quebec (Canada), and one outside Canada, in Paris (France). This inclusive process provided a better understanding of the issues citizens are most concerned about. And it also helped build trust in AI deployment because it involved an open and transparent process that allowed for the integration of citizen feedback.⁹

Ensuring that all can benefit from the deployment of AI in key sectors

The use of AI is expected to generate global productivity gains of 0.8-1.4% per year by 2065 (compared to 0.3%, between 1850 and 1910, for the invention of the steam engine). AI's contribution to the global economy is expected to reach USD 15.7 trillion as a result, as if 10 countries the size of Canada appeared on the planet. For example, AI on its own has the potential to generate more than £600bn in the U.K. economy by 2035¹⁰. The impacts of AI will be felt in all key sectors. AI could especially help address the needs of excluded or underserved people or groups (e.g. low-income workers, immigrants). For example:

- Using AI can help accelerate treatment in developing countries and other places where advanced medical expertise is scarce. Advances made by Montreal's Imagia will make it possible to use AI for inferring genetic information from regular imaging techniques. In patients living with lung or liver cancers, this could replace the use of invasive diagnostic surgical procedures often accessible only in urban centers. Using AI to gain diagnostic information from readily available imaging technology will accelerate diagnosis and access to treatment even patients living in some of Canada's most remote areas.¹¹
- The "relationship between transportation and social mobility is stronger than that between mobility and several other factors, like crime, elementary-school test scores or the percentage of two-parent families in a community."¹² By enabling the implementation of cheap and efficient on-demand transportation services, AI will help millions of low-income people save time and money to go to work or appointments.¹³
- Students all over the world can learn a second language like English more quickly with AI-powered services (like Duolingo's) that tailor lessons to their particular needs than with regular courses¹⁴. Robots like SoftBank Robotics's NAO can also play a role in engaging and teaching children with special needs, such as kids with autism spectrum disorder¹⁵
- Up to 40% of global crops are lost every year because of badly managed pests and diseases (and the incapacity of most growers to access the services of agricultural consultants), but the world's small farmers can now use mobile services like Germany's Plantix to upload pictures of affected crops, automatically diagnose problems and get recommendations on how to deal with a particular infestation.¹⁶ AI will help deal with the fact that climate change is damaging arable land and putting pressure on the global food supply¹⁷. In fact, AI is already in use in some Middle East and African countries to support water management or help monitor droughts¹⁸.
- AI will also serve to empower people that are or could be victims of exclusion. For example, Gartner estimates that AI-powered technologies like voice recognition, text-to-speech generation and autonomous cars could give 350m people with disabilities the possibility to enter the workplace over the next decade¹⁹.

Draft – For discussion purposes only; does not represent the views of the G7 or its members.

Some groups or countries may end up not seeing these gains because they will not have access to promising AI tools (e.g. only rich countries or hospitals may be able to use the latest AI for early detection of disease). The challenge will be to ensure that processes are developed and sufficient resources are available for the deployment of applications that will benefit excluded groups and societies.

Governments need to maximize this positive role by driving broad strategic engagement across departments. For example, the U.K. recently launched three new institutions—the Office for Artificial Intelligence, the Centre for Data Ethics and Innovation, and the AI Council—that will draw in expertise from academia and industry to ensure the potential of AI is exploited in all key sectors. It also founded the Institute for Ethical AI in Education, which will investigate how data and AI can be designed and deployed ethically within that sector. Researchers and entrepreneurs can also play a role in making sure AI benefits the whole of society. In Canada, MILA, one of the world’s leading AI research centres, recently ran a 6-week “AI for Social Good Lab” to teach technical concepts in AI (and business topics) with the goal of solving important social problems.

Deploying AI without creating or deepening divides between groups or countries

The growth of AI could have highly positive impacts in all sectors, but even well designed AI tools will have negative effects that must be mitigated to ensure the long-term economic and social inclusion of the world’s population.

On the employment front, computers and machines are now often performing labour intensive tasks with more efficiency and effectiveness than humans. The OECD believes the most dire forecasts of job losses resulting from automation will not materialize, but nevertheless it adds that there is still potential trouble ahead: “the risks are of ‘further polarisation of the labour market’ between highly paid workers and other jobs that may be ‘relatively low paid and not particularly interesting.’”²⁰

The consequences of this transformation will be felt strongly in the Global South, predicts Kai-Fu Lee. “With manufacturing and services increasingly done by intelligent machines located in the AI superpowers, developing countries will lose the one competitive edge that their predecessors used to kick-start development²¹.” Indeed, a 2016 study conducted by the International Labour Organization found that more than half the textile factory jobs in five Southeast Asian countries were “at high risk of automation²².” But the rise of AI may also have a strong effect in some parts of industrialized nations. Among other things, writes the Brookfield Institute, it “is likely that smaller cities and towns, heavily reliant on one industry or employer, will experience [AI-induced] automation differently from either larger or more diversified economies, regardless of the proportion of work activities with the potential to be automated. These areas may have a lesser ability to reabsorb displaced labour²³.”

Vigorous actions will need to be taken, at all levels, to avoid that countries, regions or people are left behind because of the implementation of AI. Vinod Kumar of Tata Communications explains that AI has the power to remould and better develop countries from within, by allowing tech talent to thrive: “In India, the private sector has created 4,200 AI start-ups that ‘are redefining business models in technology for healthcare,

Draft – For discussion purposes only; does not represent the views of the G7 or its members.

education, climate change, ecommerce and so forth. They will make very pinpointed and disruptive investments that will shape the economy of the future, and improve the lives of the poor in India and beyond.”²⁴ In Korea, leaders of a firm taking part in a conference called ReframeWork told how their employees were invited to find opportunities to automate the processes they were involved in and were guaranteed to keep their job after that work was done²⁵.

Governments will also play a role. France’s AI Strategy contains the recommendation that a “permanent structure” with the capacity to anticipate and experiment to modernize labour and professional training policies be created²⁶. Japan’s Advisory Board on Artificial Intelligence and Human Society states that to fight the AI divide, one of the most important issues will be to create a “space for dialogue among people with different visions and ideas and to consider common, fundamental social values²⁷.” For its part, the European Union has started to draft AI ethics guidelines that will address issues such as the future of work, fairness, safety, security and social inclusion²⁸.

Designing, developing and deploying AI in a way that does not lead to the exclusion of people or groups

There are high risks that because of the way they are designed and developed, some AI tools can lead to the exclusion of people or groups that are often already fragile. As Safiya Umoja Noble stated, “algorithms and AI applications are prone to copy human shortcomings, and replicate biased decisions that reinforce ongoing inequalities²⁹.” For example, Pulitzer Prize winner ProPublica found that COMPAS, an instrument used to guide sentencing, was more likely to incorrectly flag black defendants as being at high risk of recidivism than white defendants, but less likely to erroneously label blacks than whites as being at low risk of recidivism³⁰. Similarly, researchers have established that programs that predict when and where crimes will occur (e.g. PredPol) can lead the police to disproportionately target minorities and low-income neighbourhoods, because these programs basically reproduce, confirm and can even reinforce the prejudices that previously existed in the databases that were used to build them³¹.

In a different vein, a study has shown that because of the way Google’s online advertising platform worked a few years ago, men were exposed to ads for high-income jobs 5 times more frequently than women³². For her part, Cathy O’Neil has described how the instrument created by AI experts to help Xerox find good call centre workers at first discriminated against poorer applicants, since (1) poorer people tend to live in poorer neighbourhoods (2) poorer neighbourhoods tend to be located further away from workplaces than more affluent ones and (3) people with longer commutes tend to change job faster, of course, an unwanted trait³³.

Another risk is that the combined use of big data, connected technologies and AI tools like face recognition or sentiment analysis to make predictions and conclusions lead to the exclusion of certain citizens, workers or consumers. For example, the *Guardian* mentions “in the midst of the great deluge of personal data that comes from our online lives, there is every sign of these methods being massively extended. [...] Facebook patented a system of credit rating that would consider the financial histories of people’s friends. Opaque innovations known as e-scores are used by increasing numbers of companies to target their marketing, while such outfits as the already infamous

Draft – For discussion purposes only; does not represent the views of the G7 or its members.

Cambridge Analytica trawl people’s online activities so as to precisely target political messaging³⁴.”

Private sector players have started to develop products to mitigate risks of that type. For example, IBM has recently launched software to detect bias and explain decision-making in AI models³⁵, whereas Google released a special tool, “What-If,” to assess algorithmic fairness³⁶. But much more is needed.

For example, the Toronto Declaration enunciates that end users should be more closely involved than they are often during the design and implementation phases of AI tools as that would “help ensure that systems ‘are created and used in ways that respect rights—particularly the rights of marginalised groups who are vulnerable to discrimination.’ It also insists that States should ensure that ‘existing measures to prevent against discrimination and other rights harms are updated to take into account and address the risks posed by machine learning technologies³⁷.’ AI Ethics Researcher Abhishek Gupta mentions we should fold AI systems ‘into the purview of existing legislative and regulatory bodies to allow for certification of the quality of outputs that can be expected from these systems.’ In its forthcoming AI strategy, Germany mentions the importance of ‘ensuring the transparency, traceability and verifiability of [AI] systems, to more effectively protect against distortions, discrimination, manipulation or other improper use, especially when deploying algorithm-based forecasting and decision-making systems³⁸.’ Finally, the U.K. estimates in its AI Review that one of the ‘main ways to address these kinds of biases is to ensure that developers are drawn from diverse gender, ethnic and socio-economic backgrounds, and are aware of, and adhere to, ethical codes of conduct’ since a ‘diverse group of programmers reduces the risk of bias embedding into the algorithm and enables a fairer and higher quality output³⁹.’ It also believes we should increase accountability (especially when it comes to public algorithms) by assessing mechanisms that ensure people are still involved in decision-making, and providing those whose lives are impacted with a rationale for an algorithmic decision.

Using AI to detect potential victims of exclusion

Finally, AI offers opportunities to support diversity by revealing and disarming the biases that exist in our systems. Indeed, ‘AIs can be developed that can detect biases, both in new AI-supported functions, but also in existing, historical systems that still influence decision-making in different sectors. AI can address the challenges faced by individuals because of unconscious bias, by bringing these to the surface more effectively than has been done in the past⁴⁰.’

For example, researchers have used deep neural networks to automatically assess organizational diversity in the 500 largest companies listed in the 2016 Forbes Global 2000 rankings⁴¹. Moreover, machine learning applications can already predict, with 90% precision in some cases, if a citizen is at risk of being chronically excluded from society. This allows social workers to further study these cases and could help prevent social exclusion. Machine learning might also make it possible to answer questions like: ‘Will generational transmission of poverty occur in this family?’ or ‘How much economic aid is needed to integrate this person into society?⁴²’). Researchers have also started to study how civil society could use AI to change existing political paradigms and enable strong political

Draft – For discussion purposes only; does not represent the views of the G7 or its members.

participation beyond elections, for example by making it easier for NGOs to analyse governmental data and make recommendations to correct existing problems⁴³.

Possible Collective G7 Actions

G7 countries should try to answer the questions above collectively and work together to tackle the hard technical, social and human challenges to which they correspond. This could be done inside a new global forum like the one that France and Canada proposed to create earlier this year. Indeed, in June 2018, ‘reaffirming the G7 Innovation Ministers’ Statement on Artificial Intelligence adopted in Montreal on March 28, 2018, Canada and France announced their wish to promote ‘a vision of human-centric artificial intelligence grounded in human rights, inclusion, diversity, innovation and economic growth through the creation of an international study group⁴⁴’.

¹ See The Algorithm from MIT Technology Review <newsletters@technologyreview.com, November 9, 2018.

² See <https://www.statista.com/statistics/607716/worldwide-artificial-intelligence-market-revenues/>.

³ See <https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/>.

⁴ See <https://www.nytimes.com/2017/06/24/opinion/sunday/artificial-intelligence-economic-inequality.html?action=click&module=RelatedLinks&pgtype=Article>.

⁵ See <https://www.bbc.co.uk/mediacentre/latestnews/2017/microbit-first-year>.

⁶ See <http://ai-4-all.org/>.

⁷ See https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf.

⁸ See <https://www.elementai.com/fr/news/2018/element-ai-opens-london-office>.

⁹ See <https://www.montrealdeclaration-responsibleai.com>.

¹⁰ See https://ai.quebec/wp-content/uploads/sites/2/2018/06/AI-Strategy_EN-ACJ-19-juin-v8.pdf.

¹¹ Interview with Imagia.

¹² See <https://www.nytimes.com/2015/05/07/upshot/transportation-emerges-as-crucial-to-escaping-poverty.html>

¹³ See <https://cra.org/ccc/wp-content/uploads/sites/2/2016/06/Pascal-van-Hentenryck-AI-slides.pdf>

¹⁴ See http://static.duolingo.com/s3/DuolingoReport_Final.pdf.

¹⁵ See <http://news.mit.edu/2018/personalized-deep-learning-equips-robots-autism-therapy-0627>.

¹⁶ See <https://www.fastcompany.com/40468146/machine-learning-helps-small-farmers-identify-plant-pests-and-diseases>.

¹⁷ See <https://startups.co.uk/startups-awards/peoples-champion-finalist-2017-wefarm/>.

¹⁸ See <https://www.idrc.ca/en/stories/artificial-intelligence-and-human-development>.

¹⁹ See <https://www.gartner.com/smarterwithgartner/from-disability-to-superability/>.

²⁰ See <https://www.axios.com/automation-jobs-future-retraining-workers-e9fea761-18c3-49bb-834f-6d25819e1a1c.html>

²¹ See Kai-Fu Lee, *AI Superpowers*, 2018.

²² See <https://foreignpolicy.com/2018/07/16/closing-the-factory-doors-manufacturing-economy-automation-jobs-developing/>

²³ See https://brookfieldinstitute.ca/wp-content/uploads/RP_BrookfieldInstitute_Automation-Across-the-Nation.pdf.

²⁴ See <https://www.telegraph.co.uk/business/tata-communications/artificial-intelligence-in-developing-countries/>

²⁵ Discussion with Marc-Étienne Ouimette, Element AI.

²⁶ See https://www.aiforhumanity.fr/pdfs/MissionVillani_Summary_ENG.pdf.

²⁷ See http://www8.cao.go.jp/cstp/tyousakai/ai/summary/aisociety_en.pdf.

²⁸ See <https://ec.europa.eu/transparency/regdoc/rep/1/2018/EN/COM-2018-237-1-EN-MAIN-PART-1.PDF>.

²⁹ See Safiya Umoja Noble, *Algorithms of Oppression, How Search Engines Reinforce Racism*, 2018.

³⁰ See <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>.

³¹ See <https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x>.

³² See <https://www.cmu.edu/news/stories/archives/2015/july/online-ads-research.html>.

³³ See <https://www.theguardian.com/science/2016/sep/01/how-algorithms-rule-our-working-lives>. In the end, Xerox « sacrificed a bit of efficiency for fairness » by removing the « time to commute » indicator from its model ».

³⁴ See <https://www.theguardian.com/commentisfree/2018/mar/05/algorithms-rate-credit-scores-finances-data>.

³⁵ See <https://techcrunch.com/2018/09/19/ibm-launches-cloud-tool-to-detect-ai-bias-and-explain-automated-decisions/>.

³⁶ See <https://ai.googleblog.com/2018/09/the-what-if-tool-code-free-probing-of.html>.

³⁷ See https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf.

³⁸ See https://www.bmbf.de/files/180718%20Eckpunkte_KI-Strategie%20final%20Layout.pdf.

³⁹ See

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf.

⁴⁰ See

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf.

⁴¹ See <https://arxiv.org/abs/1707.02353>.

⁴² See <https://link.springer.com/article/10.1007/s10115-018-1230-x>.

⁴³ See www.researchgate.net/publication/326281752_Empowering_Political_Participation_through_Artificial_Intelligence.

⁴⁴ See http://international.gc.ca/world-monde/international_relations-relations_internationales/europe/2018-06-07-france_ai-ia_france.aspx?lang=eng.