Horizon Scanning Series

The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing

Global Governance

This input paper was prepared by Andrea Renda

Suggested Citation

Renda, A (2018). Global Governance. Input paper for the Horizon Scanning Project “The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing” on behalf of the Australian Council of Learned Academies, www.acola.org.

The views and opinions expressed in this report are those of the author and do not necessarily reflect the opinions of ACOLA.

AI AND THE FUTURE OF GLOBAL GOVERNANCE

Andrea Renda*

There are many ways in which the evolution and uptake of Artificial Intelligence (AI) will affect and even disrupt the global order in the next decade. Importantly, governments can make choices today, which will determine whether AI will eventually become, in the words of Stephen Hawking, either the best thing, or the worst, ever to happen to humanity"[1]. Emmanuel Macron echoed this view in a recent interview, in which he argued that "AI will raise a lot of issues in ethics, in politics, it will question our democracy and our collective preferences"[2]. Similar, even more worrying concerns were voiced by twenty-six academics in a recent report on the security implications of malicious use of AI by rogue states, criminals, and terrorists[3]. Below, I briefly discuss the current technological landscape, and then focus on political stability, security, trade and AI global governance.

1 An evolving technological landscape

In order to fully appraise the magnitude of AI's prospective impact on democracy and global governance, it is important to look at it as a full "stack", comprising i.a. computing capacity, availability of advanced, high-capacity, widespread fixed and mobile connectivity, the IoT and (where available) IoV layer[4], AI-powered platforms, applications, services and content; and AI-augmented human beings. This layered ecosystem will evolve at breath-taking speed in the coming decades, ultimately creating a data-rich layer surrounding and dominating human bodies and permeating all aspects of social interactions, from work to leisure, sex, and war. In this new, information-rich environment, and with an expected 100 connected devices per human being by 2035, humans will most likely need personalized AI to perform daily tasks[5].

Importantly, the competitive race between superpowers will focus on all these layers, with countries adopting different strategies, and different levels of government involvement in the development and implementation of technological solutions. This is visible already today. At the lower layer of the ecosystem, the race for supercomputers is leading to important breakthroughs, such as the evolution of computing capacity into parallel computing, in which different processing units perform different functions (CPUs the main tasks, GPU the graphics, TPU machine learning and other forms of AI). An even bigger breakthrough is expected when quantum computers will reach a suitable capacity (i.e., number of qubits): Google's recent decision to open-source its quantum computing platform *Cirq* is a good demonstration of the importance of securing competitive advantage early on in this rising domain[6]. Quantum and other forms of chips (e.g. biological, neuromorphic) are expected to reach the capacity of the human brain (approximately 85 billion neurons) by 2025, and then skyrocket afterwards to unknown frontiers[7]. The country that will win this race will achieve key advantages, especially in cryptography, with applications mostly in scientific research, complex optimization issues and, inevitably, cyberwarfare. It should therefore come to no surprise that China has intensified its efforts in filing patent applications for quantum cryptography, an area in which it massively dwarfs other countries today[8]. On the side of quantum computing *per se*, the US still seem to preserve a competitive advantage, mostly due to its private sector giants such as Google, IBM and Microsoft, and emerging mavericks such as Rigetti.

Even more strategic is the race for conquering the blossoming AI-enabled platform and application layers. The geopolitical relevance of this race cannot be overstated. Suffice it to recall what Vladimir Putin recently

---

* Professor of Digital Innovation, College of Europe, Bruges (Belgium), Senior Research Fellow, Centre for European Policy Studies, Brussels (Belgium).

[1] See Hawking's original interview with the BBC in 2014, at https://www.bbc.co.uk/news/technology-30290540.

[2] https://www.wired.com/story/emmanuel-macron-talks-to-wired-about-frances-ai-strategy/

[3] See Brundage, M. et al. (2018), https://img1.wsimg.com/blobby/go/3d82daa4-97fe-4096-9c6b-376b92c619de/downloads/1c6q2kc4v_50335.pdf.

[4] IoT stands for Internet of Things; whereas IoV stands for Internet of Value, a concept which encompasses current Distributed Ledger technologies (DLTs).

[5] https://community.arm.com/cfs-file/__key/telligent-evolution-components-attachments/01-1996-00-00-00-01-30-09/Arm-_2D00_-The-route-to-a-trillion-devices-_2D00_-June-2017.pdf.

[6] https://ai.googleblog.com/2018/07/announcing-cirq-open-source-framework.html

[7] See i.a. https://www.nature.com/articles/d41586-017-07523-y.

[8] See The Economist (2017). *Technology Quarterly: Here, There and Everywhere. Quantum technology is beginning to come into its own*, 2017, at https://www.economist.com/news/essays/21717782-quantum-technology-beginning-come-its-own. See also Purdy, M. and P. Dougherty (2017), Why Artificial Intelligence is the Future of Growth, Accenture/Frontier Economics Report. At https://www.accenture.com/t20170927T080049Z__w__/us-en/_acnmedia/PDF-33/Accenture-Why-AI-is-the-Future-of-Growth.PDFla=en.

said in a speech, that AI "is the future, not only for Russia but for all humankind. Whoever becomes the leader in this sphere will become the ruler of the world"[9]; and to look at the Chinese ambitious AI plan, which shows the ambition of becoming the global leader in AI innovation by 2030. Such bold initiatives have later triggered plans in many developed countries in the world, including the UK, France and the European Union. While the former two nations appear determined to invest in industrial applications to boost their global competitiveness, the latter appear more concerned with the need for "responsible AI", which falls in line with key ethical and legal principles, thus avoiding a race to the bottom in AI development. The U.S. notably does not have an official AI strategy, but exhibits massive levels of public and private R&D funding, and private initiatives to raise awareness on the need for human-compatible, responsible AI.

An important aspect of this emerging global competition is the incentive that AI creates for governments to control human behavior. Even competitiveness-oriented plans such as the Chinese and Russian ones often aim at increasing the speed of AI development by securing massive data availability for machine training: such data is often collected through very intrusive technological means, such as facial and body recognition, or even social credit scores. With a projected one trillion connected devices by 2035, the unconstrained collection of data may lead to unpredictable changes in the way governments relate to citizens, and the latter organize their social life. To be sure, in these countries the AI race may end up sacrificing the protection of personal data on the altar of faster, more capable machines. An "unintended" outcome that these governments may not despise.

The race is hectic, and the risks are high: not surprisingly, AI was recently portrayed as likely to increase the likelihood of a nuclear war in the coming two decades[10]. Making the right choices is indeed essential for the harmonious development of AI at the domestic and international level. Below, specific issues are addressed, related to both domestic and international aspects.

2 International and national political stability: anarchy v. autarchy

It is difficult to disentangle the already emerging trends in global governance from the additional effects that will be negated by the evolution and diffusion of AI technologies. Already now, China is on its way to become the most powerful global economy, with the United States currently stepping back from the proactive, almost uncontested leadership that they have enjoyed over the past decades. This is reflected in many domains of global governance, including climate policy, trade policy, and to some extent even the G7/G20. China rises as a would-be global leader even in environmental policy, and possibly also as a contributor to global peace and stability (at least in tis neighboring regions), and as a would-be leader in general-purpose new technological developments such as AI. This, coupled with the booming demography of countries like India and Nigeria, provides a first idea of the terrain on which AI will develop.

The international political order will be affected heavily by this transition. In particular, it is reasonable to expect that due to the data-hungry nature of current AI applications (mostly based on machine learning), and the pervasive nature of such applications, the AI stack will be considered as "critical infrastructure", i.e. essential to national stability in the near future, most likely within the next ten years in many developed countries. The explosion of the Internet of Things and the massive generation of data-driven, AI-powered applications that run key critical infrastructure such as energy grids, Internet pipelines, the food chain, the ATM network, hospital logistics and care delivery will gradually lead countries to try to protect the AI stack as a domestic asset. The risk of foreign "intrusion" into the data architecture, already existing today (suffice it to thin about the Russian meddling in US elections), will gradually become an existential risk for governments. Thus, a temptation to invoke so-called "AI sovereignty" or "AI autarchy" will emerge, just like sovereignty-related sentiments and reactions were elicited by the Snowden revelation related to NSA mass surveillance, especially in countries like Germany and France; and the threat of Russian or Chinese meddling into elections triggered reactions in the U.S., Italy, the UK and also recently in Australia[11]. "AI sovereignty" will be even more needed in the age of quantum supremacy, given the need to avoid that advanced in cryptography provide hostile nations with important strategic advantages in global intelligence. Despite the inherently global nature of technologies like the Internet and AI, such tendency may emerge both in non-democratic countries, and in democratic ones.

[9] https://www.youtube.com/watch?v=lHd7s3I3Zb4
[10] See the report by RAND Corporation, at https://www.rand.org/pubs/perspectives/PE296.html
[11] https://www.foreignaffairs.com/articles/australia/2018-07-26/australias-fight-against-chinese-political-interference?cid=nlc-fa_fatoday-20180726.

At the domestic level, stability will be undermined by two main trends. The first is massive job automation of work, which risks leaving entire layers of the population unemployed. Independently of what the net impact of automation will be in the end (all sorts of predictions are being made), stating that there will be no disruption is in fact impossible. Governments will initially try to address this issue with increased reliance on universal basic income schemes, "robo-taxes" or similar policies. But it is unclear whether individual well-being (typically fostered by the fact of being employed, not just by economic security[12]) and social cohesion will be materially helped by these initiatives. Inequality will increasingly go beyond money availability: unequal access to education, to political life, to high-quality services will create a risk of political disruption at home. The enormous potential of AI to reduce the cost of delivery of public services may lead the more disadvantaged parts of the population to be served by "junk AI": for example, for poorer people cheap bots may replace general practitioners, small claims judges, insurance brokers, etc.

The second development is related to the manipulation of public opinion through so-called "deep fakes" and new forms of disinformation campaigns, which make AI a threat to democracy. More specifically: while it can be expected that AI-powered real-time fact-checking will dilute the possibility for post-truth political narratives, the power of AI-enabled disinformation will equally increase.

Much of this prospective impact depends on the choices governments will make to enable the diffusion of responsible, ethical AI nested in sustainable development; by efforts to create a global AI community and governance, with common rules; and by national policies aimed at accompanying job automation with reskilling of the workforce. In this respect, the efforts made by the European Union, the UK, France and sub-national governments like Quebec are to be observed with cautious optimism. In the case of the EU, a new AI strategy was launched on 25 April 2018, and a High-Level Expert group is now leading the so-called AI Alliance towards the definition of an AI Code of Ethics by early 2019, as well as the formulation of recommendations for policy and investment, in what will become a pan-European strategy for competitiveness in (and hopefully sustainable development of) Artificial Intelligence. Needless to say, such developments will only be sustainable if coupled with ways to sustain competition with less ethically oriented, possibly cheaper forms of AI: thereby another good reason for AI autarchy and "sovereignty".

3 AI and global threats: friend and foe

As all dual technologies, AI has been inevitably associated with both positive and negative potential. In particular, some commentators and scientists classified AI as an existential risk to humanity[13]; whereas others observed that AI can make catastrophic events such as a nuclear war more likely[14]. Some of the emerging risks caused by the malicious use of artificial intelligence appear as the natural continuation of existing trends. Fake news will become "deep fakes" facilitated by Generative Adversarial Networks (GANs); phishing scams will become more sophisticated; and AI-enabled cyberattacks may become more difficult to anticipate due to the enhanced use of (unsupervised) machine learning. AAn authoritative report collectively published by several institutes in February 2018 argued that "the costs of attacks may be lowered by the scalable use of AI systems to complete tasks that would ordinarily require human labor, intelligence and expertise. A natural effect would be to expand the set of actors who can carry out particular attacks, the rate at which they can carry out these attacks, and the set of potential targets".[15] Max Tegmark's book *Life 3.0* notes the concern of UC Berkeley computer scientist Stuart Russell, who worries that the biggest winners from an AI arms race would be "small rogue states and non-state actors such as terrorists" who can access these weapons through the black market. Tegmark writes that after they are "mass-produced, small AI-powered killer drones are likely to cost little more than a smartphone." Would-be assassins could simply "upload their target's photo and address into the killer drone: it can then fly to the destination, identify and eliminate the person, and self-destruct to ensure that nobody knows who was responsible."

These risks are already sufficient to generate reactions on the government side such as the restructuring of cybersecurity and cyber-resilience plans, with the creation of pervasive, diffuse networks of data collection points, coupled with the centralization of processing power into high performance computers. However, new risks will also emerge: for example, the explosion in the number of connected devices and progress on miniaturization will lead to possible body hacking, which may concentrate on wearables and implants; the

[12] See i.a. Stam, K. et al. (2015), *Employment status and subjective well-being: the role of the social norm to work*, Work, employment and society 1–25. Available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.905.6476&rep=rep1&type=pdf.

[13] For example, Elon Musk has branded artificial intelligence "a fundamental existential risk for human civilisation".

[14]

[15] https://img1.wsimg.com/blobby/go/3d82daa4-97fe-4096-9c6b-376b92c619de/downloads/1c6q2kc4v_50335.pdf

use of self-driving cars may make road traffic a favorite target for cyber-attackers; and so-called "swarming attacks" by distributed networks of autonomous robotic systems cooperating at machine speed will become possible. At the same time, as a dual use technology, AI is also a response to other emerging risks, such as pandemics and bioterrorism. For example, companies like AIME (AI in Medical Epidemiology) have created a Dengue Outbreak Prediction platform; and in South Korea scientists have been able to train AI to detect anthrax at fast speeds.

4 An enemy of international trade?

The past few years have been characterized by a growing cross-border flow of data, but also by the rise of so-called "digital protectionism" measures, such as i.a. restrictions to foreign direct investment in high-tech markets; and measures adopted at the national level to prevent data from being stored outside the national territory. Several domestic regulations around the world are working to regulate the transfer of data abroad. In May 2017, 34 countries had reportedly adopted legislation to protect the international transfer of data: personal data (including health) and public data but also tax, accounting and financial data. Such can be detrimental to growth in many jurisdictions, especially when measures affect also non-personal data (which are anyway difficult to disentangle from personally identifiable data in many instances). Ferracane (2017) establishes a simple taxonomy, ranging from the absence of restrictions on cross-border flows to the total ban on transfer, including the obligation to store or process data locally (via cloud servers, Internet servers and data centers)[16]. As a matter of fact, such restrictions include also localization requirements of computer installations, disclosure of computer source codes (computer operating systems), the discriminatory treatment of digital products (music, video, software, e-books) transmitted electronically, and regulatory divergences on network neutrality *et similia*.

The more the data architecture and infrastructure are considered to be critical resources, the more likely that these measures proliferate in the future. Establishing trust even between traditional friendly allies such as the U.S. and the EU has proven to be an uphill battle, and there is no reason to expect that this will change significantly, and in favor of more free trade, in the years to come. If anything, the temptation to launch ambitious industrial policy strategies at the national or regional level will be hard to resist: for example, the EU is currently considering the creation of a "CERN for AI" and even an "AIRBUS for AI", which would be the catalyst of an "all European" strategy on Artificial Intelligence. Even when free-trade-oriented proposals are adopted, such as the EU "free flow of data" regulation, national security remains the possible cause of an exception, likely to be increasingly invoked[17].

More generally, countries will try to impose their standards and rules on AI by limiting possibilities for foreign companies. Aronson (2018) explains that AI is not being leveraged by governments to create competitive advantage, and to promote local values and the local economy: countries are just beginning to figure out how best to use and to protect various types of data that are used in AI, whether proprietary, personal, public or metadata. From Merkel to Macron and Trudeau, global leaders appear more inclined to use AI to boost local values than to open up to global AI development, thereby leaving most of the profits and control to US-based or china-based tech giants. Accordingly, in the coming decade AI may not positively contribute to free trade, but rather to a more fragmented trade landscape.

5 A frustrating lack of prospective global governance agreement on AI

Based on what discussed above, ongoing calls for a widespread global governance agreement on the use of AI and related standards appears to be far from likely. However, certain uses of AI could be subject to a global moratorium or outright ban, in order to avoid that countries engage in a dangerous competitive race, with possible destructive consequences for all. This is a possible outcome for the ongoing discussions on banning autonomous weapons. Academics like Toby Walsh and initiatives like the Campaign to Stop Killer Robots have denounced the escalation of this potentially destructive race, with prototype autonomous weapons under development "in every theatre of war – in the air, on the sea, under the sea and on the land"[18]. However, even in this case, difficulties in agreeing over definitions on autonomous weapons, and on patterns of attribution in case of distributed (e.g. swarming) attacks may lead the proposed agreement to collapse.

[16] http://ecipe.org/publications/restrictions-to-cross-border-data-flows-a-taxonomy/
[17] http://europa.eu/rapid/press-release_IP-18-4227_en.htm
[18] https://www.theguardian.com/technology/2018/apr/09/killer-robots-pressure-builds-for-ban-as-governments-meet

Accordingly, while there would certainly be room for fruitful agreement in the international community, the chances that such an agreement will end up being comprehensive and effectively implemented are tiny.

A different angle to the interface between global governance and AI, which seems to hold more promise in the international community, is the incorporation of AI in the overall discussion on Sustainable Development Goals. This approach, represented in ongoing initiatives such as ITU's "AI for Good Global Summit", focuses on the uses of Ai that can help the global community achieve the SDGs. This focus was also shared and echoed by several large private companies and foundations, which profess their commitment to achieving the 2030 goals through enhances use of AI. Accordingly, the near future may hold more hope for a fruitful mainstreaming of AI in the SDG debate, rather than a global treaty or agreement on AI alone.

This also implies that autonomous weapons, cyberwarfare and possible negative effects of AI on jobs, social equality and cohesion, and the environment may not be subject to a global governance effort in the next few years. A finding that can lead to very different reactions at national level: one scenario is that countries devote enhanced efforts to seek more trust-based enforcement mechanism to control that no one is using forbidden AI-enabled weapons; an alternative scenario is one of increased protectionism, and declining level of trust between national powers. With unpredictable consequences.