FRIVÄRLD

Jean-Marc Rickli

CONTAINING EMERGING TECHNOLOGIES' IMPACT ON INTERNATIONAL SECURITY

On May 6, 2018, the European General Data Protection Regulation (GDPR), which entered into force in 2016, was transposed into all EU countries' legislation. The core of GDPR is the protection of EU citizen's fundamental right to data protection. The same month, at the technology Code Conference in California in May 2018, Mary Meeker, partner at one of Silicon Valley's top venture capital firms, Kleiner Perkins, warned European regulators that "while it's crucial to manage for unintended consequences, it's also irresponsible to stop innovation and progress, especially in a world where there are a lot of countries that are doing different things." She was, in a way, summarizing Silicon Valley's self-perception that, "technology is the answer to all our problems. Now, please get out of our way."[1] This example illustrates how differently the governance and the impact of emerging technology are perceived on either side of the Atlantic. In a time when artificial intelligence (AI) and technologies of the Fourth Industrial Revolution are achieving major breakthroughs, it is crucial to adapt global governance structures so as to accompany the beneficial uses of these technologies and to avoid their malicious uses. Despite what 'technoptimists' from Silicon Valley think, international governance of emerging tech is very much needed. Below, I will review some potentially malicious uses related to AI and then addresses the global governance of emerging technologies.

The Fourth Industrial Revolution, a term coined by Klaus Schwab, the founder of the World Economic Forum, is characterized by "a fusion of technologies that is blurring the lines between physical, digital and biological spheres."[2] While railroads, electricity, and the rise of computer technology characterized earlier industrial waves, artificial

1 Chris O'Brien (2018). "Mary Meeker's Annual Valentine to Silicon Valley Reminds us Tech Utopianism is Alive and Well," *VentureBeat*, 15 June.

2 Klaus Schwab (2016). "The Fourth Industrial Revolution: What it Means, How To Respond," *World Economic Forum,* 14 January.

This brief is published from the Transatlantic Leadership Forum that was financed by the U.S. Embassy in Sweden

Jean-Marc Rickli is the Head of Global Risk and Resilience at the Geneva Centre for Security Policy in Geneva, Switzerland. He is also a research fellow at King's College London and a senior advisor for the AI Initiative at the Future Society at Harvard Kennedy School.

intelligence, robotics, 3D printing, biotechnology, neuroscience, the internet of things and quantum computing are the building blocks of the current revolution.

Artificial intelligence, which emerged in the early 1950s, has only since the early 2010s reached profoundly disruptive potential. For instance, the amount of computing power used in AI has been increasing every 3.5 months since 2012, an increase of a factor of 300,000 at the time of writing.[3] Such growth is transformative and raises security concerns. Tesla's and Space X's CEO Elon Musk and the late theoretical physicist Professor Stephen Hawking are amongst the most famous personalities that have raised concerns about future developments of AI. A recent study by an interdisciplinary group of AI experts, philosophers and political analysts also warned against the malicious uses of artificial intelligence in three security domains: digital, physical and political.[4]

Due to limited space, it is impossible to review all potential malicious uses of AI, or to touch upon the other emerging technologies. However, we can already see areas where AI could be used in a nefarious way. AI's comparative advantage is that it can scale up at a superhuman speed any activities in which enough digital data can be used. Machine-driven communication tools coupled with videos, pictures and voice-editing algorithms are unleashing unseen ways for mass manipulation. Deepfake technology, which uses AI deep learning techniques to swap faces over, has democratised the ability to create perfect visual manipulations.[5] Voice-mimicking assistants such as Google Duplex can now reproduce anyone's voice. Generative adversarial networks (GANs), which are algorithms relying on two neural networks competing with each other, can create highly realistic forged videos of policymakers and state leaders (or anyone) making fake statements. The combination of voice and image forgery has now made any piece of media on the internet suspect. A recent study looking at the state of Deepfake development showed that more than 14,000 Deepfake videos can be found on the Internet and these videos, mostly porn videos (96% of them targeting only women), have been watched more than 134 million times.[6] This is all the more disturbing considering that first Deepfake video was created only in November 2017.

Beyond malicious uses of AI, its weaponization has become a growing matter of concern for the international community and the United Nations. Since 2014, through the UN Convention on Certain Conventional Weapons (CCW), governmental experts debate whether lethal autonomous weapons systems should be banned as well as how such weapons should be constrained so as to guarantee an appropriate level of human control over the decision to kill human beings. Developments of such weapons offer indeed frightening prospects, not just because of their killing and disruptive power but also because of the ease with which these weapons could proliferate, especially in the cyber domain, as well as used as surrogates in future warfare.[7] Turkey announced that it will deploy autonomous weaponised drones, Kargu-2, equipped with facial recognition features that could work in swarms in early 2020. The level of autonomy and the precision of facial recognition is still unclear but this illustrates a worrying trend in the weaponization of AI.[8]

3 Dario Amodei and Danny Hernandez. "AI and Compute," *OpenAI Blog*, 16 May 2018.

4 Miles Brundage, *The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation*, University of Oxford: Future of Humanity Institute, February 2018.

5 Alessandro Cauduro. "Live Deep Fakes – You Can Now Change Your Face to Someone Else's in Real Time Video Applications", *Medium*, 4 April 2018.

6 Deeptrace (2019). *The State of Deepfakes: Landscape, Threat, and Impact.* Amsterdam, Deeptrace, September.

7 Andreas Krieg and Jean-Marc Rickli (2019). *Surrogate Warfare: The Transformation of War in the Twenty-first Century*. Washington: Georgetown University Press.

8 Wolfe Franck (2019). "Companies Developing Lethal Autonomous Weapons, As Groups Seek Ban, Report Says," *Avionics International*, 2 December.

Many experts point out, however, that current threats related to AI do not stem from the prospects of any artificial general intelligence (AGI) or superintelligence that would go rogue and eradicate humanity, but from the misconceptions and malfunctions of AI applications applied to our daily life as well as from their failure to integrate with different platforms or legacy systems.[9]

Machine learning algorithms work by processing thousands (or sometimes millions) of pieces of data to be operational. The issue of data integrity and biases is an area of growing concern in algorithm development. For instance, a recent study conducted at MIT demonstrated that an algorithm trained to perform image captioning that was trained with a set of pictures depicting death would then interpret any pictures taken from a Rorschach test as a murder. Norman is, for its inventors, the "world's first psychopath AI."[10] This experiment was conducted to raise awareness about data biases. Concrete current operational problems point to the same problem: a lack of a global unified governance in AI. For instance, a recent study has counted 84 documents worldwide containing ethical guidelines or principles for AI that clustered around five principles: transparency, justice and fairness, non-maleficence, responsibility and privacy. The conclusion of this study, however, points out that there is 'substantive divergence in relation to how these principles are interpreted; why they are deemed important; what issue, domain or actors they pertain to; and how they should be implemented.'[11]

Strides have been made in terms of cooperation to combat the weaponization of AI but such initiatives have not been adequate so far to address the issue. Indeed, for instance, the UN Governmental Group of Experts on Lethal Autonomous Weapons Systems has finally come up with a list of 11 agreed principles.[12] Yet, as the Article 36 NGO, specialized in reducing harm from weapons, rightly observed, 'the experts are tasked with adopting 'consensus recommendations in relation to the clarification, consideration and development of aspects of the normative and operational framework on emerging technologies in the area of lethal autonomous weapons systems' – a mandate that leaves ample room for initiatives pointing in radically different directions.'[13] Thus, initiatives coming from the private sectors and the civil society have tried to fill this gap. The Campaign to Stop Killer Robots, which preemptively seeks to ban fully autonomous weapons, has been instrumental in raising international awareness on the moral dilemmas and dangers of artificial intelligence, encouraging wider engagement on the topic. In December 2017, the largest professional engineer's organization, IEEE, published a code of conduct, the primary goal of which is to ensure that every technologist prioritizes ethical considerations in the design and development of autonomous and intelligent systems.[14]

Increasingly, leading actors in the tech industry are recognizing the importance of ensuring the positive development of AI and have been spearheading initiatives to address the issue. Among such initiatives is the Future of Life Institute, which gained particularly high visibility in 2015 for issuing an Open Letter that gathered over 8,000

9 Missy Cumming and al. (2018). *Artificial and International Affairs: Disruption Anticipated*, Chatham House Report, June.

10 Pinar Yanardag (2018). Norman World's First Psychopath AI," *MIT Media Lab*, 1st April.

11 Anna Jobin, Marcello Ienca an Effy Vayena (2019). "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence*, 2 September, pp. 389-399.

12 UN CCW (2019). *Report of the 2019 Session of the Governmental Group of Experts on Emerging Technologies in the Field of Autonomous Weapons Systems.* 25 September.

13 Article36 (2019). "Struggling for Meaning at the CCW." Geneva, 20 November.

14 IEEE (2017). *Ethically Aligned Design: a Vision Prioritising Human Well-Being with Autonomous and Intelligent Systems,* Version 2.

signatures, on *Research Priorities for Robust and Beneficial Artificial Intelligence*. The priorities put forth in the letter and its accompanying paper include verification measures, security against unauthorized manipulation, and methods for continuous and reliable human control of AI as important areas of research.[15] A similar initiative, the Partnership on AI, is a non-governmental organization founded by a coalition of tech giants: Amazon, Google, Facebook, IBM, Microsoft and Google. The partnership aims to raise awareness of AI technologies and develop and share best practice in the research, development and fielding of AI technologies. Similarly, OpenAI, a non-profit AI research company sponsored by individuals such as Peter Thiel and Elon Musk, and by companies such as Microsoft and Amazon, seeks to build safe artificial general intelligence and ensure that AGI's benefits will be as widely and as evenly distributed as possible.

Currently, leaders in the tech industry and the scientific community, as well as think-tanks and NGOs, play the most active roles in awareness-raising and cooperation on AI. Going forward, it will be essential to increase the engagement of a large range of actors, from private and start-up companies to governments and international organizations, in order to institute a comprehensive system to safeguard the future applications of AI in our daily lives.[16]

The disruptive potential of artificial intelligence but also of the other technologies derived from the so-called Fourth Industrial Revolution as well as their accelerated rate of advancement signal that we could soon be living in an unrecognizable world. What distinguishes the technological revolution we face today from past periods of change is the degree of control humans are surrendering to machines whose decision-making processes we do not fully understand. Moreover, with the development of AI comes the risk that this incredibly powerful technology will be used for malicious purposes. The forgery of digital pictures, sounds or films is just an early example of the ways individuals might use AI for malicious purposes. The weaponization of artificial intelligence and autonomy offers new ways of fighting wars. Of course, accidents and unintended effects can also have detrimental consequences.

A global system of governance on AI that provides transparency in terms of AI applications (perhaps not in fundamental research, as some sensitive experiments might need protection) establishes norms of AI development, certification and application, and effectively monitors compliance, is therefore not merely a valuable foresight but a prerequisite in ensuring that AI is developed as a force for good. This will require the combined effort of the private and commercial sectors, academia, governments and international and non-governmental organizations. A holistic and comprehensive governance system should therefore be developed.

15 Stuart Russell, D. D. (2015). "Research Priorities for Robust and Beneficial Artificial Intelligence." *AI Magazine,* 36(4), pp. 105-114.

16 World Economic Forum (2019). *Global Technology Governance: a Multistakeholder Approach.* Geneva: WEF, October.