Regulating AI: Considerations that Apply Across Domains

Angela Kane

Contents

Abstract

Awareness that AI-based technologies have far outpaced the existing regulatory frameworks have raised challenging questions about how to set limits on the most dangerous developments (lethal autonomous weapons or surveillance bots, for instance). Under the assumption that the robotics industry cannot be relied on to regulate itself, calls for government intervention within the regulatory space—national and international—have multiplied. The various approaches to regulating AI fall into two main categories. A sectoral approach looks to identify the societal risks posed by individual technologies, so that preventive or mitigating strategies can be implemented, on the assumption that the rules applicable to AI, in say the financial industry, would be very different from those relevant to heath care providers. A cross-sectoral approach, by contrast, involves the formulation of rules (whether norms adopted by industrial consensus or laws set down by governmental authority) that, as the name implies, would have application to AI-based technologies in their generality. After surveying some domestic and international initiatives that typify the two approaches, the chapter concludes with a list of 15 recommendations to guide reflection on the promotion of societally beneficial AI.

A. Kane (✉)
Vienna Center for Disarmament and Non-Proliferation (VCDNP), Vienna, Austria
e-mail: akane@vcdnp.org

Keywords

AI · Regulation · Domestic · International · Recommendations

J. von Braun et al. (eds.), *Robotics, AI, and Humanity*, https://doi.org/10.1007/978-3-030-54173-6_21

Introduction[1]

While we think of AI as a phenomenon that has rapidly arisen over the last few years, we should remember that it was already 80 years ago that Alan Turing laid down the mathematical basis of computation. ARPANET began in 1969, the Internet Protocol in 1974, and the World Wide Web 30 years ago, in 1989. Do any of you remember when we first started to access the internet, with modems? The distinctive whirring burpy sound they made when connecting—ever so slowly—to the web? This now seems very quaint, as the improvements in speed and performance, as well as the cost reductions in memory and information technology, have made possible the enormous expansion of data that now fuels the engine of global growth.

Harnessing AI has challenges and opportunities in many areas and domains: technical, ethical, political, social, and cultural. These are accompanied by the need for accountability, algorithmic explainability, and even legal liability. If we do not understand how a system works, then it blurs lines of who can or who should be responsible for the outcome or the process of the decision. Should that be the innovator? The regulator? The operator? And how can both policymakers and the public trust technology when it is not properly understood?

These are vexing questions that have been further compounded by the rise in disclosures of data and privacy leaks, of hacking into sites containing sensitive personal information, of spoofing, of selling consumer data without consent and, to make matters worse, of concealing or delaying disclosure of such egregious violations of privacy.

The debate about these issues has become louder and more polarized; it is pitting powerful companies against governments and consumers. Scientists are weighing in—as do employees of technology companies, as we have seen with Google. Until 2015, Google's motto was "Don't be evil" but it was then changed to "Do the right thing" within its corporate code of conduct. Swarms of bots, dark posts, and fake news websites inundate the web, ricochet around chatrooms, and overwhelm the legitimate media outlets.

Let us remember just a few recent events: in the US presidential elections in 2016, Russia supported one candidate (who subsequently won) by waging a campaign with paid advertisements and fake social media accounts that contained polarizing content. Concerns also abound in China about millions of cameras deployed with face recognition software which record streams of data about citizens. In India, it was reported that the "fake news problem plagues several popular social networks" (Metha 2019) by spreading misinformation, doctored photos and videos which resulted in several cases of killing and even lynching.

More and more thoughtful questions about social platforms are being asked that do not lend themselves to easy answers. Technology companies are coming under increasing scrutiny, as they are seen to be operating without accountability. Facebook CEO Mark Zuckerberg testified in US Congress on efforts to address privacy issues and data sharing, but subsequent Facebook data leaks showed that his assurances to prevent a recurrence were hollow. Talking about regulation, he said: "My position is not that there should be no regulation. I think the real question is, as the internet becomes more important in people's lives, is what is the right regulation, not whether there should be or not" (Zuckerberg in Watson 2018).

In October 2019, responding to concerns that the social network has too much power to shape political and social issues, Zuckerberg pushed back against criticism that Facebook was not doing enough to combat hate speech, misinformation, and other offensive content, by opining that people in a democracy did not want a private company censoring the news (Wong 2019). Does free speech then allow the placing of ads with deliberate falsehoods? This question has taken added relevance, particularly in the run-up to the 2020 presidential elections in the USA where the use of social media is a prime factor in the campaign.

I will take stock of some of the efforts to address the attempts to regulate AI and technology, fully aware that the paper will outdate very quickly, as new initiatives and considerations are coming up quickly.

Curbing Lethal Autonomous Weapon Systems: An Early Effort at Regulating AI

In Wikipedia's definition, artificial general intelligence is the intelligence of a machine that can understand or learn any intellectual task that a human being can. Yet while the jury is still out whether AI will bring enormous benefits to humanity or bring possible calamity, the applications of AI abound in a variety of sectors. Deep learning algorithms are embedded already in our daily life; they are used in social media, medicine, surveillance, and determining government benefits, among others. By way of example, let me therefore look at one of the sectoral approaches, that of using AI in weapons.

In 2013, a report was published by the United Nations Special Rapporteur on extrajudicial, summary or arbitrary execution, Christof Heyns, on the use of lethal force through what he called "lethal autonomous robotics (LAR)." He approached the issue from the perspective of protection of

[1] This chapter is based on an earlier version presented at the Pontifical Academy of Sciences in May 2019 that was also placed on the website https://www.united-europe.eu/2019/05/angela-kane-regulating-ai-considerations-that-apply-across-domains/

life during war and peace and made a number of urgent recommendations to organizations, States, developers of robotic systems, and nongovernmental organizations.

Following its publication, 16 countries put the questions related to emerging—or "robotic"—technologies on the agenda of the Convention on Certain Conventional Weapons (CCW) in Geneva. The first meetings on these issues took place in 2014 and they showed that few countries had developed any policy on the matter. Thematic sessions, with significant input from AI scientists, academics, and activists, dealt with legal aspects, ethical and sociological aspects, meaningful human control over targeting and attack decisions, as well as operational and military aspects. And what Christof Heyns had called lethal autonomous robotics is now referred to as "lethal autonomous weapon systems" or LAWS. While there is no singularly accepted definition of LAWS, the term now generally covers a broad array of potential weapon systems, from fully autonomous weapons that can launch attacks without any human intervention to semiautonomous weapons that need human action to direct or execute a mission.

The first debates were conducted in an Open-Ended Working Group, in which any State could freely participate. Yet in 2016, governments decided to form a Group of Governmental Experts (GGE) to advance the issue. The crucial difference is that a GGE operates on a consensus basis, which essentially gives a veto right to any decisions or statements adopted by the GGE to any one participating State.

Twenty-nine States now openly call for a ban on these weapons. Austria, Brazil, and Chile have recently proposed a mandate to "negotiate a legally-binding instrument to ensure meaningful human control over the critical functions of weapon systems," but the prospects for such a move are slim. So far, no legally binding or political actions have been adopted by the Group due to the objections of about a dozen States: Australia, Belgium, France, Germany, Israel, Republic of Korea, Russian Federation, Spain, Sweden, Turkey, the United Kingdom, and the United States. These States argue that concrete action on LAWS is "premature" and that the Group could instead explore "potential benefits" of developing and using LAWS.

The opposing positions do not augur well for any legislative progress in the issue of LAWS. Yet the voices in favor of a total ban are getting louder and louder. Already in 2015, at one of the world's leading AI conferences, the International Joint Conference on Artificial Intelligence (IJCAI 15), an Open Letter from AI & Robotics Researchers—signed by nearly 4000 of the preeminent scientists such as Stuart Russell, Yann LeCun, Demis Hassabis, Noel Sharkey, and many many others—and over 22,000 endorsers including Stephen Hawking, Elon Musk, and Jaan Tallinn, to name just a few, warned against AI weapons development and posited that "most AI researchers have no interest in building AI weapons, and do not want others to tarnish their field by doing so" (FLI 2015).

The decision by Google to end cooperation with the US Department of Defense on Project Maven—a minor contract in financial terms—was ended in 2018 due to strong opposition by Google employees who believed that Google should not be in the business of war. UN Secretary-General Guterres, former High Commissioner for Human Rights Zeid Ra'ad Al Hussein, and Pope Francis have weighed in, calling autonomous weapons "morally repugnant" and calling for a ban (UN 2018a).

There are also parliamentary initiatives in capitals. In April 2018, for example, the Lord's Select Committee on AI challenged the UK's futuristic definitions of autonomous weapon systems as "clearly out of step" with those of the rest of the world and demanded that the UK's position be changed to align these within a few months.

Yet the Government's response was limited to one paragraph which stated that the Ministry of Defense "has no plans to change the definition of an autonomous system" and notes that the UK will actively participate in future GGE meetings in Geneva, "trying to reach agreement (on the definition and characteristics of possible LAWS) at the earliest possible stage" (UK Parliament 2018, recommendations 60–61).

Interest in other European parliaments is also high, as awareness of the issue has grown exponentially. It is the hot topic of the day.

The European Commission issued a communication in April 2018 with a blueprint for "Artificial Intelligence for Europe" (European Commission 2018). While this does not specifically refer to LAWS, it demands an appropriate ethical and legal framework based on the EU's values and in line with the Charter of Fundamental Rights of the Union.

In July 2018, the European Parliament adopted a resolution that calls for the urgent negotiation of "*an international ban on weapon systems that lack human control over the use of force.*" The resolution calls on the European Council to work towards such a ban and "*urgently develop and adopt a common position on autonomous weapon systems*" (European Parliament 2018). In September 2018, EU High Representative Federica Mogherini told the EU Parliament that "*the use of force must always abide by international law, including international humanitarian law and human rights laws. (. . .) How governments should manage the rise of AI to ensure we harness the opportunities while also addressing the threats of the digital era is one of the major strands of open debate the EU has initiated together with tech leaders*" (EEAS 2018).

The issue of lethal autonomous weapons has clearly raised the profile of legislating AI. Advocacy by civil society, especially the Campaign to Stop Killer Robots, a coalition of NGOs seeking to pre-emptively ban lethal autonomous weapons, has been instrumental in keeping the issue prominent in the media, but this single-issue focus is not easily replicable in other AI-driven technologies.

Can We Ever Hope to Regulate and Govern AI?

Artificial intelligence is a universal subject that breaks down into many variations and applications. Rather than tackling AI as a whole, it is easier to address a sector-specific AI application—like LAWS—than general AI that is broad, adaptive, and advanced as a human being across a range of cognitive tasks.

We already have a myriad of automated decision systems that are being used by public agencies, in criminal justice systems, in predictive policing, in college admissions, in hiring decisions, and many more. Are these automated decision systems appropriate? Should they be used in particularly sensitive domains? How can we fully assess the impact of these systems? Whose interests do they serve? Are they sufficiently nuanced to take into account complex social and historical contexts? Do they cause unintended consequences?

The difficulty in finding answers to these questions is the lack of transparency and information. Many of these systems operate in a black box and thus outside the scope of understanding, scrutiny and accountability. Yet algorithms are endowed with a specific structuring function, as designed by individuals. The General Data Protection Regulation (GDPR) which the European Union adopted in 2018 includes an "explainability requirement" that applies to AI, but it is not clear exactly how much.

"Can You Sue an Algorithm for Malpractice?" was the headline of a magazine article in the USA in 2019 (Forbes 2019). Clearly, algorithms are being litigated, as a 2018 report by the AI Now Institute shows (AI Now 2018a) and which has resulted already in more study and scrutiny of the use of such systems across public agencies. Several lawsuits proved that decision-making formulas were corrupt due to data entry errors and biased historical data, while aimed to produce cost savings or to streamline work without assessment how they might harm vulnerable populations. While this showed the limits of AI use in public policy, it is clear that lawsuits set precedent in law but cannot establish regulations and the rule of law.

But if the litigation shows us anything, it is that AI-driven technology has become an important issue for people and for governments. In response, we are seeing two distinct trends:

- The AI and tech industry have become a hub for ethics advisory boards and related efforts to buff their credentials in what I would call "responsible AI".
- Private organizations have been established like Partnership for AI (mission: to benefit people and society), or Open AI (mission: to ensure that artificial general intelligence benefits all of humanity).
- Academic institutions—such as New York University—have set up institutes like AI Now, a research institute examining the social implications of AI; the Massachusetts Institute of Technology (MIT) conducts a project on AI Ethics and Governance to support people and institutions who are working to steer AI in ethically conscious directions.
- Workshops and conferences with a range of tech and non-tech stakeholders are being organized to debate the scope of the challenges as well as exploring solutions.

Governments Are Stepping Up

The second trend is the increasing focus by Governments on the disruption by artificial intelligence and the search for shaping the ethics of AI. Let me mention some statements by leaders.

When Russian President Putin in 2018 said to a group of school children that "whoever controls AI, will become the ruler of the world," it made headlines. China's blueprint—issued in 2017 and called the "New Generation Artificial Intelligence Development Plan"—outlined China's strategy to become the world player in AI by 2030. The Plan barely mentions information on laws, regulations, and ethical norms, since China's authoritarian approach is less restrained by attention to values and fundamental rights as well as ethical principles such as accountability and transparency. In the 3 years since its publication, China is already starting to overtake the USA as the leader in AI.

In Europe, French President Macron in 2018 called the technological revolution that comes with AI "in fact a political revolution," and said that in shaping how AI would affect us, you have to be involved at the design stage, *and set the rules* (italics added). He committed the French government to spend Euro 1.5 billion over 5 years to support research in the field, encourage startups, and collect data that can be used, and shared, by engineers.

A French data protection agency (Commission Nationale de l'Informatique et des Libertés, CNIL) issued a 75-page report in December 2017 about the results of a public debate about AI, algorithms, ethics, and how to regulate it. The report set out six areas, which predominate the ethical dilemmas:

1. Autonomous machines taking decisions
2. Tendencies, discrimination and exclusion which are programmed, intentionally or unintentionally
3. Algorithmic profiling of people

4. Preventing data collection for machine learning
5. Challenges in selecting data of quality, quantity, and relevance
6. Human identity in the age of artificial intelligence

Recommendations made in the report primarily focus on the individual by urging enhanced information and education but also request private industry to focus on ethics by establishing ethics committees and an ethics code of conduct or an ethics charter (CNIL 2017).

In the UK, the House of Lords Select Committee on Artificial Intelligence issued a report in April 2018 with the catchy title "AI in the UK: ready, willing and able?" The report was based on extensive consultations and contains an assessment of the current state of affairs as well as numerous recommendations on living with AI, and on shaping AI (House of Lords Artificial Intelligence Committee 2018).

The 183-page report has only two paragraphs on "regulation and regulators" which state that "Blanket AI-specific regulation, at this stage, would be inappropriate. We believe that existing sector-specific regulators are best placed to consider the impact on their sectors of any subsequent regulation *which may be needed*" (emphasis added). It also urges the Government Office for AI to "ensure that the existing regulators' expertise is utilized in informing any potential regulation that may be required in the future" and foresees that "the additional burden this could place on existing regulators could be substantial," recommending adequate and sustainable funding (House of Lords Artificial Intelligence Committee 2018: 386–387). In its final paragraphs, the report refers to the preparation of ethical codes of conduct for the use of AI by "many organizations" and recommends that a cross-sectoral ethical code of conduct—suitable for implementation across public and private sector organizations—be drawn up (. . .) with a sense of urgency. "In time, the AI code could provide the basis for statutory regulation, *if and when this is determined to be necessary*" (House of Lords Artificial Intelligence Committee 2018: 420, emphasis added).

In June 2018, the Government issued a 42-page response to the House of Lords' report. As to paragraph 386 (no blanket AI-specific regulation needed), the Government agreed with the recommendation. It stated its commitment to work with businesses to "develop an agile approach to regulation that promotes innovation and the growth of new sectors, while protecting citizens and the environment" (UK Parliament 2018). It further promises horizon-scanning and identifying the areas where regulation needs to adapt to support emerging technologies such as AI and the establishment of a Centre for Data Ethics and Innovation that "will help strengthen the existing governance landscape" (UK Parliament 2018: 108). Yet the Centre—established late last year—has only an advisory function, promoting best practices and advising how Government should address potential gaps in the regulatory landscape.

Other European countries also addressed AI. Sweden published a report in May 2018 on its National Approach (a digestible 12 pages) which highlights the Government's goals to develop standards and principles for ethical, sustainable, and safe AI, and to improve digital infrastructure to leverage opportunities in AI. Finland was a bit ahead of the curve, issuing its first report on "Finland's Age of Artificial Intelligence" already in December 2017, but none of its eight proposals deal with rules and regulations.

Germany issued a 12-point strategy ("AI Made in Germany—a seal of excellence"), which focuses on making vast troves of data available to German researchers and developers, improves conditions for entrepreneurs, stops the brain drain of AI experts, and loosens or adapts regulation in certain areas. But it also heavily emphasizes the rights and advantages of AI for the citizens and underlines the ethical and legal anchoring of AI in Europe.

The European Union: "Placing the Power of AI at the Service of Human Progress"

Finally, let me focus on the European Union which in April 2018 issued "AI for Europe: Embracing Change" (European Commission 2018). This was the launch of a European Initiative on AI with the following aims:

1. Boost the EU's technological and industrial capacity and AI uptake across the economy
2. Prepare for socio-economic change
3. Ensure an appropriate ethical and legal framework

Under these three headings, ambitious plans were laid out, both in financial terms (stepping up investments) and in deliverables, with time lines until the end of 2020.

Let us not forget that the General Data Protection Regulation (GDPR) came into force the same year. While this regulation imposes a uniform data security law on all EU members, it is important to note that any company that markets good and services to EU residents, regardless of its location, is subject to the regulation. This means that GDPR is not limited to EU member states, but that it will have a global effect.

One of the deliverables was the setting up of an Independent High-Level Expert Group on Artificial Intelligence[2] which was asked to draft AI ethics guidelines and through an online framework called the European AI Alliance reached out to stakeholders and experts to contribute to this effort.

[2] Full disclosure: I was a reserve member of the High-Level Expert Group and participated in several of their meetings.

The draft ethics guidelines were issued in December 2018 and received over 500 comments, according to the EU. What resulted were the "Ethics Guidelines for Trustworthy AI," issued in April 2019, which defines trustworthy AI as follows: "*(It) has three components: (1) it should be lawful, ensuring compliance with all applicable laws and regulations (2) it should be ethical, demonstrating respect for, and ensure adherence to, ethical principles and values and (3) it should be robust, both from a technical and social perspective since, even with good intentions, AI systems can cause unintentional harm. Trustworthy AI concerns not only the trustworthiness of the AI system itself but also comprises the trustworthiness of all processes and actors that are part of the system's life cycle*."

The Guidelines then list seven essentials for achieving trustworthy AI:

1. Human agency and oversight
2. Robustness and safety
3. Privacy and data governance
4. Transparency
5. Diversity, non-discrimination, and fairness
6. Societal and environmental well-being
7. Accountability

Again, the Guidelines are currently in a pilot phase for more time to receive feedback and to ensure that they can be issued by the end of 2019 and then implemented—which is expected in 2020 (European Commission 2019). At the same time, the EU Commission wants to bring their approach to AI ethics to the global stage: "because technologies, data and algorithms know no borders." Following the G7 summit in Canada in December 2018, where AI was prominently featured, the EU wants to strengthen cooperation with other "like-minded" countries like Canada, Japan, and Singapore, but also with international organizations and initiatives like the G20 to advance the AI ethics agenda.

Before we break out the champagne in celebration of the ethics guidelines, let me mention one dissenting voice from the High-Level Group: Thomas Metzinger, Professor of Theoretical Philosophy in Germany, wrote a scathing article entitled "Ethics washing made in Europe" in which he called the Trustworthy AI story "a marketing narrative invented by industry, a bedtime story for tomorrow's customers." The narrative, he claimed is "in reality, about developing future markets and using ethics debates as elegant public decorations for a large-scale investment strategy" (Metzinger 2019). Metzinger (2019) considers that "industry organizes and cultivates ethical debates to buy time—to distract the public and to prevent or at least delay effective regulation and policy-making. And politicians like to set up ethics committees because it gives them a course of action when, given the complexities of the issues, they simply don't know what to do." Interestingly, he also mentions the use of lethal autonomous weapon systems as one of the "Red Lines," the non-negotiable ethical principles—which I outlined at the beginning of this paper.

Ethical AI—The New Corporate Buzz Phrase

I agree that the jury on the EU Ethics Guidelines is still out, but the criticism of major tech companies and academic ethics boards, especially in the USA, is very strong. Many tech companies have recently laid out ethical principles to guide their work on AI. Major companies like Microsoft, Facebook, and Axon (which makes stun guns and body cameras for police departments), all now have advisory boards on the issue. Amazon recently announced that it is helping fund research into "algorithmic fairness," and Salesforce employs an "architect" for ethical AI practice, as well as a "chief ethical and human use" officer. More examples could be cited.

Yet are these actions designed primarily to head off new government regulations? Is it a fig leaf or a positive step? "Ethical codes may deflect criticism by acknowledging that problems exist, without ceding any power to regulate or transform the way technology is developed and applied," wrote the AI Now Institute, a research group at New York University, in a 2018 report. "We have not seen strong oversight and accountability to backstop these ethical commitments" (AI Now 2018b).

The boards are also seen to mirror real-world inequality (mostly white men, very few women, few or no people of color or minorities) (see Levin 2019) or to have members who do not represent ethical values. The establishment of an ethics board by Google (actually called Advanced Technology External Advisory Council, ATEAC) lasted barely a week before it was disbanded amid great controversy.

The Google debate shows that discussing these issues in the public eye also invites public scrutiny. While I consider it positive that private industry is studying the issues and inviting views on company ethics, it is ultimately the CEO who gets to decide which suggestions on AI ethics would be incorporated into what are essentially business decisions. A company is clearly more concerned with the financial bottom line rather than sacrificing profit for ethical positions taken by an external advisory board, as there is no legal obligation to follow what are well-intentioned recommendations.

So the issue revolves around accountability, and in my view, government regulation will be needed to enforce it. Doteveryone, a UK organization (mission: Responsible Technology for a Fairer Future), issued a report entitled "Regulating for Responsible Technology" (Miller et al. 2018) which calls for a new independent regulatory body with three responsibilities:

1. Give regulators the capacity to hold technology to account.
2. Inform the public and policy-makers with robust evidence on the impacts of technology.
3. Support people to seek redress from technology-driven harms.

In addition to outlining that we currently have a "system in need of a steward," the organization also has a directory of regulation proposals in the UK to which it invites users to update (Doteveryone not dated). More surveys of such proposals might be very helpful in determining how best to go forward.

We should, however, also look at "soft law" which are substantive expectations that are not directly enforceable, as opposed to "hard law" which are legally enforceable requirements imposed by governments. As outlined by Wallach and Marchant, soft law includes voluntary programs, standards, codes of conduct, best practices, certification programs, guidelines, and statements of principles (Wallach and Marchant 2019). As an example of soft law being turned into hard law, they cite the Future of Life Institute Asilomar Principles (FLI 2017) adopted in 2017 as a soft law tool for AI governance, which have now been adopted by the State of California into its statutory law.

A Paradigm Shift Is Emerging

I believe one of the problems of the EU's High-Level Expert Group on AI is that it tries to be all-comprehensive and therefore tends towards more general and lofty declarations rather than be prescriptive in application. As I noted at the beginning of this paper, it is easier to address regulation in one aspect of AI rather than the entire gamut of applications. Let me focus on one such aspect that has started to capture attention in a major way: facial recognition and the pervasive use of cameras.

The Turing Award has been given to three preeminent computing scientists for their work on neural networks which has, inter alia, accelerated the development of face-recognition services. Yet they—together with some two dozen prominent AI researchers—have signed a letter to Amazon to stop selling its face-recognition technology (called "Rekognition") to law enforcement agencies because it is biased against women and people of color.

Facial recognition technology (FRT) has been used by government agencies, by retail industry, by Facebook with its millions of users posting photographs. In China, more than 176 million CCTV cameras are used for street monitoring and policing as well as in "cashless" stores and ATMs: where does consumer assistance start and surveillance begin?

Despite some positive aspects (reuniting missing children in India), there are major concerns about how to protect the privacy of those whose data is collected. With an industry quickly mushrooming to an estimated more than $10 billion in the next few years, alarms are beginning to sound about the lack of governmental oversight and the stealthy way it can be used to collect data on crowds of people—as we learned when it was revealed that the musician Taylor Swift had deployed FTR during her performances to root out stalkers. But is the technology only used for security?

Containing FTR is easier in Europe, where strict privacy laws are being enforced with the GDPR, but in other countries (and continents) no regulations exist. Yet even here in Europe people are warning against the "surveillance state." Looking at the increasing coverage and discussion of FTR, I am of the opinion that this will be one area of focus for regulation in the near future.

Could There Be a Role for International Organizations or Institutions?

UN Secretary-General Antonio Guterres weighed in on AI in July 2018, stating that "the scale, spread and speed of change made possible by digital technologies is unprecedented, but the current means and levels of international cooperation are unequal to the challenge (UN 2018b)." He set up a High-Level Panel on Digital Cooperation, with Melinda Gates and Jack Ma as Co-Chairs, and 18 additional members serving in their individual capacity. Their task was to submit a report by mid-2019—contributing to the broader public debate—which identified policy, research, and information gaps, and made proposals to strengthen international cooperation in the digital space.

The Panel has reached out and sought comments on their efforts from people all over the world, conducting a "global dialogue" to assist in reaching their final conclusions. Of course, it is important to bring this discussion to all member states, many of which do not have the capacity to harness new technology and lack a sophisticated understanding of the matter. It is also important for the Organization to embed this report in the universal UN values, and to consider practical ways to leverage digital technologies to achieve the Sustainable Development Goals.

The report—called "The Age of Digital Interdependence"—emphasizes the importance of fostering greater inclusivity and trust online and sets out recommendations for potential models of cooperation, yet the report is more of a summary overview of the current state of affairs rather than a model for implementation of ideas and suggestions (UN Secretary General 2019). It is vague how the report's wide-sweeping recommendations will be applied, and there appears no direct follow-up.

What is missing, in my opinion, is to take stock of existing—and emerging—normative, regulatory, and cooperative processes. I would not expect the UN to set rules and standards, but to have an inventory of the current state of affairs would be very valuable for national efforts to build on.

Past efforts by UN high-level panels have had mixed success. Despite the enormous work that goes into reports by high-ranking participants, their recommendations have at times been taken note of, politely debated—and then disappeared into a drawer without seeing implementation. Let us hope that the prominent co-chairs of this report will continue to contribute to a lively open debate and ensure that the recommendations will see further discussion and follow-up.

Summing Up: 15 Recommendations

Rapidly emerging technologies—AI and robotics in particular—present a singular challenge to regulation by governments. The technologies are owned by private industry, they advance in the blink of an eye, and they are not easily understood due to their complexity and may be obsolete by the time a government has agreed to regulate them.

This means that traditional models of government regulation cannot be applied. So if not regulation, what can be done? Here are my proposals:

1. Expand AI expertise so that it is not confined to a small number of countries or a narrow segment of the population.
2. Accept that the right decisions on AI technology will not be taken without strong input from the technologists themselves.
3. Find therefore a common language for government officials, policy-makers, and technical experts.
4. Begin dialogue so that (a) policies are informed by technical possibilities and (b) technologists/experts appreciate the requirements for policy accountability.
5. Discuss how to build a social license for AI, including new incentive structures to encourage governments and private industry to align the development and deployment of AI technologies with the public interest.
6. Focus on outcome, not process: principles, privacy protection, digital policy convergence, and differences in legal and regulatory systems and cultures between the USA, EU, and China.
7. Establish some "Red Lines"—no-go areas for AI technology, such as lethal autonomous weapon systems, AI-supported assessment of citizens by the government ("social scoring").
8. Use the strategy of "soft law" to overcome limitations and challenges of traditional government regulation for AI and robotics.
9. Discuss the challenges, costs, reliability, and limitations of the current state of art.
10. Develop strong working relationships, particularly in the defense sector, between public and private AI developers.
11. Ensure that developers and regulators pay particular attention to the question of human-machine interface.
12. Understand how different domains raise different challenges.
13. Compile a list of guidelines that already exist and see where there are gaps that need to be filled to offer more guidance on transparency, accountability and fairness of AI tools.
14. Learn from adjacent communities (cyber security, biotech, aviation) about efforts to improve safety and robustness.
15. Governments, foundations, and corporations should allocate funding to develop and deploy AI systems with humanitarian goals.

I encourage others to add to the list. What is really important here is that we come to a common understanding of what needs to be done. How do we develop international protocols on how to develop and deploy AI systems? The more people ask that question, the more debate we have on it, the closer we will get to a common approach. This is what is needed more than ever today.

References

AI Now. (2018a). *Litigating algorithms: Challenging government use of algorithmic decision systems*. AI Now Institute. Retrieved from https://ainowinstitute.org/litigatingalgorithms.pdf

AI Now. (2018b). *AI Now report 2018*. AI Now Institute. Retrieved from https://ainowinstitute.org/AI_Now_2018_Report.pdf

CNIL. (2017). *How can humans keep the upper hand? Report on the ethical matters raised by algorithms and artificial intelligence*. Retrieved from https://www.cnil.fr/en/how-can-humans-keep-upper-hand-report-ethical-matters-raised-algorithms-and-artificial-intelligence

European External Action Services (EEAS). (2018). *Autonomous weapons must remain under human control, Mogherini says at European Parliament*. Retrieved from https://eeas.europa.eu/topics/economic-relations-connectivity-innovation/50465/autonomous-weapons-must-remain-under-human-control-mogherini-says-european-parliament_en

European Commission. (2018). *Artificial intelligence for Europe*. Communication from the commission to the European Parliament, the European council, the council, the European economic and social committee and the committee of the regions, Brussel, 25 April 2018. Retrieved from https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51625

European Commission. (2019). *Ethics guidelines for trustworthy AI*. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

European Parliament. (2018). *Resolution of 12 September 2018 on autonomous weapon systems (2018/2752(RSP))*. Retrieved from https://www.europarl.europa.eu/doceo/document/TA-8-2018-0341_EN.html

The Future of Life Institute (FLI). (2015). *Autonomous weapons: An open letter from AI & robotics researchers*. Retrieved from https://futureoflife.org/open-letter-autonomous-weapons

The Future of Life Institute (FLI). (2017). *Asilomar AI principles*. Retrieved from https://futureoflife.org/ai-principles/

Forbes. (2019). *Can you sue an algorithm for malpractice?* Interview with W. Nicholson Price. Forbes insights, 11 February 2019. Retrieved from https://www.forbes.com/sites/insights-intelai/2019/02/11/can-you-sue-an-algorithm-for-malpractice/

House of Lords Artificial Intelligence Committee. (2018). *AI in the UK: Ready, willing and able*. Retrieved from https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

Levin, S. (2019). *'Bias deep inside the code': The problem with AI 'ethics' in Silicon Valley*. The Guardian, 29 March 2019. Retrieved from https://www.theguardian.com/technology/2019/mar/28/big-tech-ai-ethics-boards-prejudice

Metha, I. (2019). *It's not just WhatsApp—India's fake news problem plagues several popular social networks*. The Next Web, 29 January 2019. Retrieved from https://thenextweb.com/in/2019/01/29/its-not-just-whatsapp-indias-fake-news-problem-plagues-several-popular-social-networks/

Metzinger, T. (2019). *Ethics washing made in Europe*. Der Tagesspiegel, 8 April 2019. Retrieved from https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html

Miller, C., Ohrvik-Stott, J., & Coldicutt, R. (2018). *Regulating for responsible technology: Capacity, evidence and redress: A new system for a fairer future*. London: Doteveryone. Retrieved from https://doteveryone.org.uk/project/regulating-for-responsible-technology/.

UK Parliament. (2018). *Government response to house of lords artificial intelligence select committee's report on AI in the UK: Ready, willing and able?* Retrieved from https://www.parliament.uk/documents/lords-committees/Artificial-Intelligence/AI-Government-Response.pdf

United Nations (UN). (2018a). *Machines with power, discretion to take human life politically unacceptable, morally repugnant, secretary-general tells Lisbon 'Web Summit'*. Press release, 5 November 2018. Retrieved from https://www.un.org/press/en/2018/sgsm19332.doc.htm

United Nations (UN). (2018b). *Secretary-general appoints high-level panel on digital cooperation*. Press release, 12 July 2018. Retrieved from https://www.un.org/press/en/2018/sga1817.doc.htm

UN Secretary General. (2019). *The age of digital interdependence—Report of the high-level panel on digital cooperation*. Retrieved from https://digitalcooperation.org/report

Wallach, W., & Marchant, G. (2019). Toward the agile and comprehensive international governance of AI and robotics. *Proceedings of the IEEE, 107*(3), 505–508.

Watson, C. (2018). *The key moments from Mark Zuckerberg's testimony to Congress*. The Guardian, 11 April 2018. Retrieved from https://www.theguardian.com/technology/2018/apr/11/mark-zuckerbergs-testimony-to-congress-the-key-moments

Wong, Q. (2019). *Facebook CEO Mark Zuckerberg pushes back against claims of anti-conservative censorship*. Cnet, 18 October 2019. Retrieved from https://www.cnet.com/news/facebook-ceo-mark-zuckerberg-pushes-back-against-claims-of-conservative-censorship/