GENEVA ACADEMY
Académie de droit international humanitaire et de droits humains
Academy of International Humanitarian Law and Human Rights

RESEARCH BRIEF

HUMAN RIGHTS AND THE GOVERNANCE OF ARTIFICIAL INTELLIGENCE

KEY MESSAGES

- Artificial intelligence (AI) is bound to enable innovation in the decades to come, so much so that some say it has become the new electricity.[1] However, if that truly is the case, then policymakers, business and civil society must understand what the opportunities and challenges are before they turn the switch on. AI enthusiasts forecast that such technologies could improve societal well-being, increase productivity and even provide solutions for global climate and health crises. AI could also help fight human rights abuses. Nonetheless, AI presents a variety of challenges that can profoundly affect the respect for and protection of human rights.

- Recently, a profusion of initiatives from a variety of actors spanning from the technology industry to international and regional organizations, academia and civil society, have focused on establishing ethical frameworks for the design and implementation of AI solutions. While these valuable initiatives propose to identify core ethical principles applicable to AI, ethics is only one aspect to be taken into consideration. International Human Rights Law (IHRL) is equally, if not more important.

- Stakeholders from the private and public sectors, international organizations and civil society should move beyond the calls for more regulation of AI. Regulation is certainly needed, in particular concerning data protection and privacy. Nonetheless, new models of governance, placed alongside regulatory frameworks and existing human rights instruments, are also needed. This research brief identifies two additional avenues to regulation: public procurement and standardization.

FEBRUARY 2020 I ANA BEDUSCHI

THE GENEVA ACADEMY A JOINT CENTER OF
THE GRADUATE INSTITUTE GENEVA
INSTITUT DE HAUTES ÉTUDES INTERNATIONALES ET DU DÉVELOPPEMENT
GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES
UNIVERSITÉ DE GENÈVE
FACULTY OF LAW

INTRODUCTION

AI has the potential to revolutionize the way both the public and the private sectors operate. AI technologies currently power virtual assistants on smart devices, provide fraud alerts for banking applications and help improve health diagnostics.[2] AI solutions are also increasingly used in sectors such as law enforcement, judicial decision-making, border security, international migration management and the military.[3]

To date, there is no single agreed definition of AI. In general, AI can be understood as the 'systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals'.[4] Simply put, AI involves the use of techniques allowing machines to come closer to some aspects of human cognition.[5] Machine learning is one of these techniques, by which machines are trained to perform tasks that are generally associated with human intelligence such as natural language processing.[6] Deep learning, a subset of machine learning, is also increasingly being relied on for image and face recognition.[7] Machines learn from vast amounts of data using algorithms (i.e. sets of instructions used to solve problems). AI algorithms can analyse data, find patterns, make inferences and predict behaviour at a level and speed greatly surpassing human capabilities. Deep learning structures algorithms into layers to create an artificial neural network, enabling machines to learn and make decisions on their own.[8]

Artificial intelligence (AI) is bound to enable innovation in the decades to come, so much so that some say that it has become the new electricity.[9] However, if that truly is the case, then policymakers, business and civil society must understand what the opportunities and challenges are before they turn the switch on.

OPPORTUNITIES AND RISKS

AI enthusiasts forecast that these technologies could improve well-being, increase productivity and even provide solutions for global climate and health crises.[10] AI could also help fight human rights abuses. For example, AI algorithms can sift through large amounts of data to establish patterns and identify financial transactions that may be indicative of human trafficking networks.[11]

Nonetheless, AI presents a variety of challenges that can profoundly affect the respect and protection of human rights. Firstly, human biases, even unconscious ones, may permeate the design and development of AI systems.[12] For instance, the choice of datasets at the start of the machine-learning process can contain important biases. When, for example, AI algorithms are trained with predominantly male datasets such as men's CVs, they will replicate and give more weight to predominantly male characteristics when used for hiring employees.[13] Inherent biases may lead to

1 S. Lynch, 'Andrew Ng: Why AI Is the New Electricity' (2017), *Insights*, Stanford Business, 11 March 2017, https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity (last accessed 13 January 2020).

2 See notably, M. Chui, M. Harryson, J. Manyika, R. Roberts, R. Chung, A. van Heteren and P. Nel, *Notes From the AI Frontier: Applying AI for Social Good*, McKinsey Global Institute, 2018; P. A. Keane and E. J. Topol, 'With an Eye to AI and Autonomous Diagnosis', 1 *NPJ Digital Medicine* 40 (2018); T. Panch, H. Mattie and L. A. Celli, 'The "Inconvenient Truth" About AI in Healthcare', 2 *NPJ Digital Medicine* 77 (2019).

3 See notably, A. G. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, New York University Press, 2017; M. Hamilton, 'The Biased Algorithm: Evidence of Disparate Impact on Hispanics', 56 *American Criminal Law Review* 1553 (2018); A. Beduschi, 'The Big Data of International Migration: Opportunities and Challenges for States Under International Human Rights Law', 49 *Georgetown Journal of International Law* 981 (2018); P. Molnar, 'New Technologies in Migration: Human Rights Impacts', 61 Forced Migration Review 7 (2019); A. Beduschi, 'International Migration Management in the Age of Artificial Intelligence', (2020) volume 8, pages 1-2, https://doi.org/10.1093/migration/mnaa003; *Migration Studies*; Organisation for Economic Co-operation and Development (OECD), *Artificial Intelligence in Society*, 2019; House of Lords Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing and Able?*, HL Paper 100, 2018; N. Goussac, 'Safety Net or Tangled Web: Legal Reviews of AI in Weapons and War-Fighting' *Humanitarian Law & Policy*, 18 April 2019, https://blogs.icrc.org/law-and-policy/2019/04/18/safety-net-tangled-web-legal-reviews-ai-weapons-war-fighting/ (last accessed 20 February 2020); L. McGregor, 'The Need for Clear Governance Frameworks on Predictive Algorithms in Military Settings', Humanitarian Law & Policy, 28 March 2019, https://blogs.icrc.org/law-and-policy/2019/03/28/need-clear-governance-frameworks-predictive-algorithms-military-settings/ (last accessed 20 February 2020).

4 European Commission, *Artificial Intelligence for Europe*, COM(2018) 237 final, 25 April 2018, §1.

5 R. Calo, 'Artificial Intelligence Policy: A Primer and Roadmap', 51 *UC Davis Law Review* (2017) 404.

6 P. A. Flach, *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*, Cambridge University Press, 2012; N. J. Nilsson, *Principles of Artificial Intelligence*, Morgan Kaufmann, 2014; W. Ertel, I*ntroduction to Artificial Intelligence*, 2nd edn, Springer, 2018.

7 Y. LeCun, Y. Bengio and G. Hinton, 'Deep Learning', 521 *Nature* (2015).

8 Ibid.

9 Lynch, 'Andrew Ng', supra fn 1.

10 See notably, J. Snow, 'How Artificial Intelligence Can Tackle Climate Change', *National Geographic* , 18 July 2019, https://www.nationalgeographic.com/environment/2019/07/artificial-intelligence-climate-change/ (last accessed 13 January 2020); Expert Panel, Forbes Technology Council, *Forbes*, '15 Social Challenges AI Could Help Solve', 3 September 2019, https://www.forbes.com/sites/forbestechcouncil/2019/09/03/15-social-challenges-ai-could-help-solve/#107507693533 (last accessed 13 January 2020); A. Gray, '5 Global Problems That AI Could Help Us Solve', World Economic Forum, 7 February 2017, https://www.weforum.org/agenda/2017/02/5-global-problems-that-ai-could-help-us-solve/ (last accessed 13 January 2020).

11 Inter-Agency Coordination Group against Trafficking in Persons, 'Human Trafficking and Technology: Trends, Challenges and Opportunities', Issue Brief 7, 2019, https://www.un.org/sexualviolenceinconflict/wp-content/uploads/2019/07/report/human-trafficking-and-technology-trends-challenges-and-opportunities/Human-trafficking-and-technology-trends-challenges-and-opportunities-WEB...-1.pdf (last accessed 13 January 2020).

12 B. Friedman and H. Nissenbaum, 'Bias in Computer Systems', 14 *ACM Transactions on Information Systems* 3 (1996); V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*, St Martin's Press, 2018; S. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York University Press, 2018.

13 J. Dastin, 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women', *Reuters*, 10 October 2018, https://www.reuters.com/

unlawful discrimination based on protected characteristics, including race, gender and sexual orientation. Therefore, risks of discrimination, including indirect and intersectional cases of discrimination, in AI algorithms, must be averted or mitigated.[14]

Secondly, as current AI applications are heavily data-driven, they particularly impact data protection and privacy rights. Machine-learning algorithms 'learn' from a variety of data sources, including publicly available social media data such as videos, photos, text and audio files that individuals voluntarily upload to online platforms and applications. These types of data may contain personal information or lead to unveiling sensitive characteristics such as racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, sexual orientation and health data, which are all protected by law.[15] Moreover, due to the increasingly sophisticated ways in which online platforms and companies track online behaviour and individuals' digital footprints, AI algorithms can make inferences about behaviour, including relating to their political opinions, religion, state of health or sexual orientation.[16] Such algorithms are commonly used in the retail sector, to provide customers with personalized advertisements and purchasing suggestions based on their browsing history.[17]

Thirdly, the 'black box' nature of many AI algorithms causes problems for their explainability and auditability.[18] This is of particular concern for deep-learning models based on artificial neural networks. In these models, a machine learns and makes decisions on its own, while humans cannot explain the exact process through which the machine has reached a decision or produced an output. Computer and data scientists have since developed models and techniques to facilitate explainability and auditing of decisions made by AI algorithms, but as of now, these are neither fully functional nor widely accepted.[19] Accordingly, an algorithm may commit errors, and based on such errors, adopt a decision or present an output. Humans using such algorithms to inform their decision-making would not be able to know that there were errors in the algorithmic decision-making or output. This has obvious serious consequences for the protection of human rights. For instance, if the algorithm in question was used to predict the risk of a prisoner reoffending in the context of judicial decisions on parole,[20] undetected errors could unduly deprive the prisoner of their liberty, or conversely they could put the lives and safety of others at risk by supporting the release of a dangerous individual.

MAKING HUMAN RIGHTS RELEVANT TO THE GOVERNANCE OF AI

Against this background, it is important to reiterate that 'the same rights that people have offline must also be protected online',[21] including in the context of AI technologies. These include, in particular, the right to privacy,[22] the right to freedom of expression[23] and freedom of assembly and association,[24] and the guarantees of non-discrimination[25] and due process.[26] Human rights should thus be a key aspect of the governance of AI and not just an afterthought.

Recently, a profusion of initiatives from a variety of actors spanning from the technology industry to international and regional organizations, academia and civil society, have focused on establishing ethical frameworks for the design and implementation of AI solutions.[27] While these valuable initiatives propose to identify core ethical principles applicable to AI, ethics is only one aspect to be taken into

article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

14 See M. Mann and T. Matzner, 'Challenging Algorithmic Profiling: The Limits of Data Protection and Anti-Discrimination in Responding to Emergent Discrimination', 6 *Big Data & Society* 2 (2019).

15 See e.g., Art 9, General Data Protection Regulation (GDPR).

16 See S. Wachter and B. Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI', 2019 *Columbia Business Law Review* 2 (2019).

17 See R. Calo, 'Digital Market Manipulation' 82 *George Washington Law Review* 4 (2014).

18 See F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, 2016.

19 See e.g., Google's Explainable AI: https://cloud.google.com/blog/products/ai-machine-learning/google-cloud-ai-explanations-to-increase-fairness-responsibility-and-trust (last accessed 13 January 2020). See also, A. Rai, 'Explainable AI: From Black Box to Glass Box' , 48 *Journal of the Academy of Marketing Science* (2020).

20 J. Larson, S. Mattu, L. Kirchner and J. Angwin, 'How We Analyzed the COMPAS Recidivism Algorithm', *Pro Publica*, 23 May 2016, www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm; Hamilton, 'The Biased Algorithm', supra fn 2.

21 UNGA Res 68/167, 21 January 2014, §2; See also Human Rights Council (HRC), The Promotion, Protection and Enjoyment of Human Rights on the Internet, UN Doc A/HRC/20/L.13, 29 June 2012; HRC, The Promotion, Protection and Enjoyment of Human Rights on the Internet, UN Doc A/HRC/32/L.20, 27 June 2016; M. N. Schmitt (ed), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, Cambridge University Press, 2017, p 179.

22 Art 12, Universal Declaration of Human Rights (UDHR); Art 17, International Covenant on Civil and Political Rights (ICCPR); Art 8, European Convention on Human Rights (ECHR); Art 11, American Convention on Human Rights (ACHR).

23 Art 19, UDHR; Art 19, ICCPR; Art 10, ECHR; Art 13, ACHR; Art 9, African Charter on Human and Peoples' Rights (ACHPR).

24 Art 20, UDHR; Arts 21–22, ICCPR; Art 11, ECHR; Arts 15–16, ACHR; Arts 10–11, ACHPR.

25 Art 2, UDHR; Art 26, ICCPR; Art 14, ECHR; Art 1, ACHR; Art 2, ACHPR.

26 Art 10, UDHR; Art 14, ICCPR; Art 6, ECHR; Art 8, ACHR; Art 7, ACHPR.

27 For a general overview of these initiatives, see H. Hilligoss and J. Fjeld, 'Introducing the Principled Artificial Intelligence Project', https://cyber.harvard.edu/story/2019-06/introducing-principled-artificial-intelligence-project (last accessed 20 February 2020).

consideration. IHRL is equally important in this field. It provides a legally binding framework for dealing with human rights violations.[28] The private sector, including technology companies, can build on the existing United Nations (UN) Guiding Principles on Business and Human Rights to guide the development of new technological advances in AI.[29]

A NEED FOR COORDINATED EFFORTS

The multitude of initiatives on AI is indicative of one of the field's main challenges: most stakeholders in this area tend to operate in silos, without overall coordination of their activities and outputs. They produce multiple reports, guidelines, blueprints and statements of principles which often fail to reach beyond sectoral and specialized audiences.

International organizations such as the UN are already active in this field. For instance, the Secretary-General's High-level Panel on Digital Cooperation has recently provided important recommendations on digital technologies.[30] Nonetheless, more needs to be done to bring all the relevant stakeholders together, coordinating their efforts to tackle the challenges posed by AI. Within the UN system, the Office of the High Commissioner for Human Rights (OHCHR) is an obvious choice to lead the way, as it holds a specific mandate on human rights. The UN Global Pulse is also an important player as it carries out remarkable work in the fields of data science and AI. The Working Group on Business and Human Rights could also contribute its expertise in reaching out to the private sector. Together with other relevant stakeholders, it should work towards ensuring that human rights are firmly embedded into the design, development and deployment of AI systems across the globe. By doing so, it would act on the High-level Panel on Digital Cooperation's recommendations 3A–3C outlined in a report from 2019.[31]

TWO COMPLEMENTARY PATHWAYS

To break the existing silos, stakeholders from the private and public sectors, international organizations and civil society should move beyond the calls for more regulation of AI. Regulation is certainly needed, in particular concerning data protection and privacy.[32] Nonetheless, new models of governance, placed alongside existing regulatory frameworks and human rights instruments, are also needed. This research brief identifies two additional avenues to regulation: public procurement and standardization. Although regulation at domestic and regional levels remains a privileged method, complementary mechanisms can also include human rights principles and rules, thus shaping the future of AI.

PUBLIC PROCUREMENT

Public procurement is the process by which public bodies and authorities, including government and local authorities, purchase goods and services from businesses. Public procurement can be a powerful tool to ensure respect for human rights.[33] It is submitted that it can play a significant role also with regard to human rights in the AI sphere. Public bodies and authorities should require that suppliers respect human rights while designing, developing and deploying AI technologies that they intend to supply. Such procurement policies can enhance compliance with human rights rules, standards and principles. They can help prevent human rights abuses commonly associated with business supply chains such as modern slavery, child labour and human trafficking. They can also help tackle the lack of algorithmic fairness and accountability.

In particular, public procurement calls should contain specific clauses requiring that AI technologies are assessed to avoid any forms of discrimination, including direct, indirect and intersectional discrimination. That would require human rights impact assessments to become an indispensable element of any business practices, at least of those intending to supply public bodies and authorities. Emerging good practice on corporate human rights due diligence could also pave the way to better practices in the area of AI and new technologies.

STANDARDIZATION

As it is common practice with the Internet, AI would benefit from a set of widely agreed protocols based on

28 See L. McGregor, D. Murray and V. Ng, 'International Human Rights Law as a Framework for Algorithmic Accountability', 68 *International & Comparative Law Quarterly* 309 (2019).

29 HRC, Protect, Respect and Remedy: A Framework for Business and Human Rights, Report of the Special Representative of the Secretary-General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises, John Ruggie, UN doc A/HRC/8/5, 7 April 2008.

30 UN Secretary-General's High-level Panel on Digital Cooperation, T*he Age of Digital Interdependence*, 2019, https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf (last accessed 13 January 2020).

31 Ibid.

32 See notably, California Consumer Privacy Act 2018; European Commission, Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications), COM/2017/010 final.

33 O. Martin-Ortega and C. Methven O'Brien (eds), *Public Procurement and Human Rights: Opportunities, Risks and Dilemmas for the State as Buyer*, Edward Elgar, 2019.

technical standards. Although the Internet has developed in a much different way than AI and relies on networks to function,[34] the analogy with Internet protocols remains relevant as an illustration of the influence technical standards can have in this area. For the Internet, technical standards are set forth by two main leading organizations, the Internet Engineering Task Force (IETF) and the Internet Architecture Board (IAB).[35] They are the basis for defining the Internet protocols, i.e. the sets of rules or standards determining how data is presented and delivered on the Internet. Internet protocols based on technical standards set the framework in which the Internet continues to develop, constituting an important governing force.

Due to the societal issues posed by AI technologies, AI protocols should be based on technical standards incorporating human rights rules and principles. These standards should be set forth by a collective body representing the different sectors of society such as industry, states, civil society, international organizations and academia. It should build on the work undertaken by different organizations and groups, including the International Telecommunication Union's (ITU) work on standardization of AI, the UN Secretary-General's High-level Panel on Digital Cooperation and the European Union's (EU) High-Level Expert Group on Artificial Intelligence. Although challenging, a truly global collective body should encompass representatives from leading western and non-western powers active in the AI field, including the United States, EU, Russia and China.

Such an endeavour should strive to find common ground on how to translate human rights principles and rules including the rights to privacy, freedom of information, freedom of expression, non-discrimination and due process, to feasible technical standards. For example, technical solutions could take inspiration from the existing privacy and data protection by design protocols.[36]

Moreover, teaching the basics of human rights law to software engineers and other practitioners responsible for innovation in the AI field could contribute to improving compliance with legal and ethical requirements.[37] Scenario-based learning informed by real-life incidents complemented with detailed legal analysis of human rights rules and principles in plain language could be used to enhance the design and implementation of responsible AI.

CONCLUSIONS AND RECOMMENDATIONS

Most stakeholders tend to operate in silos, without overall coordination of their activities and outputs. The UN should take the lead and bring these stakeholders together, coordinating their efforts to tackle the challenges posed by AI. Together, they should work towards ensuring that human rights are firmly embedded into the design, development and deployment of AI systems across the globe.

As technologies evolve, new models of governance are crucially needed. Human rights should occupy a prominent place in the governance of AI. Besides regulation, public procurement and standardization should also include human rights principles and rules, thus shaping the future of AI. Public bodies and authorities should require that suppliers respect human rights while designing, developing and deploying AI technologies that they intend to supply.

Finally, AI protocols should be based on technical standards incorporating human rights rules and principles. These standards should be set forth by a collective body with global reach and representing the different sectors of society including industry, states, civil society, international organizations and academia.

ABOUT THE AUTHOR

Dr Ana Beduschi is an Associate Professor of Law at the University of Exeter. Her research and teaching focus on international human rights law, technology (including big data and artificial intelligence), digital law, as well as international migration and refugee law. This research brief is the result of the research she carried out during a three-month fellowship at the Geneva Academy.

34 B. M. Leiner, V. G. Cerf, D. D. Clark, R. E. Kahn, L. Kleinrock, D. C. Lynch, J. Postel, L. G. Roberts and S. Wolff, 'A Brief History of the Internet' 39 *ACM SIGCOMM Computer Communication Review* 5 (2009).

35 See A. Rachovitsa, 'Rethinking Privacy Online and Human Rights: The Internet's Standardisation Bodies as the Guardians of Privacy Online in the Face of Mass Surveillance', European Society of International Law Conference Paper Series 5/2016 (2016).

36 Ibid.

37 A. Beduschi, 'Technology Dominates Our Lives – That's Why We Should Teach Human Rights Law to Software Engineers', The Conversation, 26 September 2018, https://theconversation.com/technology-dominates-our-lives-thats-why-we-should-teachhuman-rights-law-to-software-engineers-102530 (last accessed 13 January 2020).

THE GENEVA ACADEMY

The Geneva Academy provides post-graduate education, conducts academic legal research and policy studies, and organizes training courses and expert meetings. We concentrate on branches of international law that relate to situations of armed conflict, protracted violence, and protection of human rights.

The Geneva Academy
of International Humanitarian Law
and Human Rights

Villa Moynier
Rue de Lausanne 120B
CP 1063 - 1211 Geneva 1 - Switzerland
Phone: +41 (22) 908 44 83
Email: info@geneva-academy.ch
www.geneva-academy.ch