

The New Regulation of the European Union on Artificial Intelligence

Fuzzy Ethics Diffuse into Domestic Law and Sideline International Law

Thomas Burri

I. INTRODUCTION

In the conventional picture, international law emanates from treaties states conclude or customs they observe. States comply with binding international law and ensure compliance in the domestic context. In this picture, states in a ‘top-top’ process agree on the law before it trickles down to the domestic legal order where it is implemented. Norms made in other ways are considered ‘soft’, which implies that they provide mere guidance but are technically not binding, or irrelevant to international law.

Obviously, there is room for nuance in the conventional take on international law and its sources. Soft law, for instance, can acquire authority that comes close to binding character.¹ It can also serve to interpret binding law that would otherwise remain ambiguous.² However, traditional international law ignores that law is also created outside of its formal processes. Norms can notably consolidate independently from the will of states in speedy, subcutaneous processes. Norms can diffuse subliminally across the world into municipal laws which incorporate and make them binding domestically. In this informal process, international law enters the stage late, if at all. It can only retrace the law that has already been locked in domestically. This informal process resembles ‘bottom-up international law’,³ though its character is more ‘bottom to bottom’ and ‘transnational’. The process shall be referred to as ‘norm diffusion’ in this chapter. It is illustrated through the creation of norms governing Artificial Intelligence (AI).

The informal process of law creation described above is far from ubiquitous. It can be hard to trace, for when international law codifies or crystallizes ‘new’ norms, it tends to obscure their origin in previous processes of law creation. It is also messy, for it does not adhere to the hierarchies that distinguish conventional international law. Even more so, it is worth discussing norm diffusion to complement the picture of international law and its sources.

The present chapter could have examined norm diffusion in the current global public health crisis. It seems that in the COVID-19 pandemic, behavioural norms informed by scientific

¹ See the treatment accorded to the ICJ, *Legal Consequences of the Separation of the Chagos Archipelago from Mauritius in 1965*, *Advisory Opinion* [2019] ICJ Rep 95, in *Dispute concerning Delimitation of the Maritime Boundary between Mauritius and Maldives in the Indian Ocean*, no. 28 (Mauritius/Maldives) (Preliminary Objections) ITLOS (2021) para. 203; see our discussion in T Burri and J Trinidad, ‘Introductory note’ (2021) 60(6) *International Legal Materials* 969–1037.

² Article 32 Vienna Convention on the Law of the Treaties, 1155 UNTS 331 (engl.) 27 March 1969.

³ J Koven Levit, ‘Bottom-Up International Lawmaking: Reflections on the New Haven School of International Law’ (2007) 32 *The Yale Journal of International Law* 393–420.

expertise take shape rapidly, diffuse globally, and are incorporated into domestic law. In contrast, international lawyers are only now beginning to discuss a more suitable legal framework. However, rather than engaging with the ongoing chaotic normative process in public health, this chapter discusses a more mature and traceable occurrence of norm diffusion, namely that of the regulation of AI. The European Commission's long-awaited proposal from April 2021 for a regulation on AI marks the perfect occasion to illustrate the diffusion of AI norms.

This chapter proceeds in three steps. First, it examines the creation of ethical norms designed to govern AI (Section II). Second, it investigates the diffusion of such norms into domestic law (Section III). This section examines the European Commission's recent legislative proposal to show how it absorbs ethical norms on AI. This examination likewise sheds light on the substance of AI norms. Section III could also be read on its own, in other words, without regard to international law-making, if one wished to learn only about the origins and the substance of the European Union regulation in the offing. Section IV then discusses how the process of norm diffusion described in Sections II and III sidelines international law. Section V concludes and offers an outlook.

II. THE CREATION OF ETHICAL NORMS ON AI

The creation of ethical norms governing AI has taken many forms over a short period of time. It began with robotics. Roughly 50 years ago, *Isaac Asimov's* science fiction showed how ambiguous certain ethical axioms were when applied to intelligent robots.⁴ Since then, robotics has made so much progress that scientists have begun to take an interest in ethical principles for robotics. Such principles, which were prominently enunciated in the United Kingdom in 2010, addressed the potential harm caused by robots, responsibility for damage, fundamental rights in the context of robotics, and several other topics, including safety/security, deception, and transparency.⁵ The same or similar aspects turned out to be relevant for AI after it had re-awakened from hibernation. Two initiatives were significant in this regard, namely the launch of the One Hundred Year Study on Artificial Intelligence at Stanford University in 2014⁶ and an

⁴ See A Winfield, 'An Updated Round Up of Ethical Principles of Robotics and AI' (*Alan Winfield's Web Log*, 18 April 2019) <https://alanwinfield.blogspot.com/2019/04/an-updated-round-up-of-ethical.html>: '1. A robot may not injure a human being or, through inaction, allow a human being to come to harm. 2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law. 3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.' The work of the present author has benefitted tremendously from Winfield's collation of ethics principles on AI in his blog at a time when it was not yet easy to assemble the various sets of ethics principles. For the primary source of Asimov's principles, see e.g. I Asimov, *The Caves of Steel* (1954) and I Asimov, *The Naked Sun* (1957); for a discussion of Asimov's principles about fifty years after Asimov had begun writing about them, see RR Murphy and DD Woods, 'Beyond Asimov: The Three Laws of Responsible Robotics' (2009) July/August 2019 *IEEE Intelligent Systems* 14–20.

⁵ Drafted in the context of the Engineering and Physical Sciences Research Council and the Arts and Humanities Research Council (United Kingdom) in 2010, but published only in M Boden and others, 'Principles of Robotics: Regulating Robots in the Real World' (2017) 29 *Connection Science* (2) 124–129; see also A Winfield, 'Roboethics – for Humans' (2011) 17 May 2011 *The New Scientist* 32–33. Before that, ethicists and philosophers had already discussed robotics in various perspectives, see e.g. R Sparrow, 'Killer Robots' (2007) 24 *Journal of Applied Philosophy* (1) 62–77, RC Arkin, *Governing Lethal Behavior in Autonomous Robots* (2009); PW Singer, *Wired for War: The Robotics Revolution and Conflict in the Twenty-First Century* (2009); W Wallach and C Allen, *Moral Machines: Teaching Robots Right from Wrong* (2009).

⁶ See E Horvitz, *One Hundred Year Study on Artificial Intelligence: Reflections and Framing* (2014) <https://ainoo.stanford.edu/reflections-and-framing> (hereafter Horvitz, 'One Hundred Year Study') also for the roots of this study (on p 1).

Open Letter⁷ signed by researchers and entrepreneurs in 2015.⁸ Both initiatives sought to guide research toward beneficial and robust AI.⁹ In their wake, the IEEE, an organization of professional engineers, in 2015 embarked on a broad public initiative aimed at pinning down the ethics of autonomous systems;¹⁰ a group of AI professionals gathered to generate the Asilomar principles for AI, which were published in 2017¹¹; and an association of experts put forward ethical principles for algorithms and programming.¹² This push to establish ethical norms occurred in

⁷ Future of Life Institute, 'An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence' (*Future of Life Institute*) <http://futureoflife.org/ai-open-letter/> (hereafter 'Open Letter'); another important moment before the Open Letter was a newspaper article: S Hawking and others, 'Transcendence Looks at the Implications of Artificial Intelligence – But Are We Taking AI Seriously Enough?' *The Independent* (1 May 2014) www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-9313474.html.

⁸ Several research groups had addressed the law and ethics of robots in the meanwhile: see C Leroux and others, 'Suggestion for a Green Paper on Legal Issues in Robotics' (31 December 2012) www.researchgate.net/publication/310167745_A_green_paper_on_legal_issues_in_robotics; E Palmerini and others, 'Guidelines on Regulating Robotics' (*Robo Law*, 22 September 2014) www.robotlaw.eu/RoboLaw_files/documents/robotlaw_d6.2_guidelinesregulatingrobotics_20140922.pdf; other authors previously had prepared the ground, notably P Lin, K Abney, and GA Bekey (eds), *Robot Ethics: The Ethical and Social Implications of Robotics* (2012); U Pagallo, *The Law of Robots: Crimes, Contracts, Torts* (2013); N Bostrom, *Superintelligence: Paths, Dangers, Strategies* (2014) (hereafter Bostrom, 'Superintelligence'); JF Weaver, *Robots Are People Too: How Siri, Google Car, and Artificial Intelligence Will Force Us to Change Our Laws* (2014); M Anderson and S Anderson Leigh, 'Towards Ensuring Ethical Behaviour from Autonomous Systems: A Case-Supported Principle-Based Paradigm' (2015) 42 *Industrial Robot: An International Journal* (4) 324–331.

⁹ In the 100 Year Study, law and ethics figured prominently as a research topic (Horvitz, 'One Hundred Year Study' (n 6) topics 6 and 7), while the Open Letter (n 7) included a research agenda parts of which were 'law' and 'ethics'.

¹⁰ The first version of 'Ethically Aligned Design' was made public in 2016: Institute of Electrical and Electronics Engineers (IEEE), 'Ethically Aligned Design, The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems' (13 December 2016) http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf; meanwhile, a first edition has become available: Institute of Electrical and Electronics Engineer (IEEE), 'Ethically Aligned Design, The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems' (2019) <https://ethicsinaction.ieee.org>; in the following, reference is made to the latter, the first edition (hereafter, IEEE, 'Ethically Aligned Design'). It contains a section on high-level 'general principles' which address human rights, well-being, data agency, effectiveness, transparency, accountability, awareness of misuse, and competence. Other sections of the Charter discuss classical ethics, well-being, affective computing, personal data and individual agency, methods to guide ethical research and design, sustainable development, embedding values, policy, and law. The last section on the 'law' focuses on fostering trust in autonomous and intelligent systems and the legal status of such systems. For full disclosure, the present author co-authored the section on law of Ethically Aligned Design.

¹¹ Future of Life Institute, 'Asilomar AI Principles' (*Future of Life Institute*, 2017) <https://futureoflife.org/ai-principles/> (hereafter Future of Life Institute, 'Asilomar AI Principles'). The Asilomar principles address AI under three themes, namely 'research', 'ethics and values', and 'longer term issues'. Several sub-topics are grouped under each theme, viz. goal, funding, science-policy link, culture, race avoidance (under 'research'); safety, failure transparency, judicial transparency, responsibility, value alignment, human values, personal privacy, liberty and privacy, shared benefit, shared prosperity, human control, non-subversion, arms race (under 'ethics and values'); and capability caution, importance, risks, recursive self-improvement, and common good (under 'longer term issues').

¹² Association for Computing Machinery US Public Policy Council (USACM), 'Statement on Algorithmic Transparency and Accountability' (USACM, 12 January 2017) www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf (hereafter USACM, 'Algorithmic Transparency'); the principles are part of a broader code of ethics: Association for Computing Machinery Committee on Professional Ethics, 'ACM Code of Ethics and Professional Conduct' (*ACM Ethics*, 22 June 2018) <https://ethics.acm.org>. Summed up, the principles are the following: 1. Be aware of bias; 2. Enable questioning and redress; 3. If you use algorithms, you are responsible even if not able to explain; 4. Produce explanations; 5. Describe the data collection process, while access may be restricted; 6. Record to enable audits; 7. Rigorously validate your model and make the test public. Compare also with the principles a professional organization outside of the anglophone sphere published relatively early: Japanese Society for Artificial Intelligence, 'The Japanese Society for Artificial Intelligence Ethical Guidelines' (2017) <http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf> (hereafter Japanese Society for AI, 'Guidelines') in summary: 1. Contribute to humanity, respect human rights and diversity, eliminate threats to safety; 2. Abide by the law, do not use AI to harm others, directly or indirectly; 3. Respect privacy; 4. AI as a resource is to be used fairly and equally by humanity, avoid discrimination and inequality; 5. Be sure to maintain AI safe and under control; provide users with

lockstep with the significant technological advances in AI,¹³ and it is against this background that it must be understood.

In parallel, a discussion began to take shape within the Convention on Certain Conventional Weapons (CCW)¹⁴ in Geneva. This discussion soon shifted its focus to the use of force by means of autonomous systems.¹⁵ It notably zeroed in on physically embodied weapons systems – a highly specialized type of robot – and refrained from considering disembodied weapons, sometimes called cyberweapons.¹⁶ The focus on embodiment¹⁷ had the effect of keeping AI out of the limelight in Geneva for a long time.¹⁸ As a broader consequence, the international law

appropriate and sufficient information; 6. Act with integrity and so that society can trust you; 7. Verify performance and impact of AI, warn if necessary, prevent misuse; whistle blowers shall not be punished; 8. Improve society's understanding of AI, maintain consistent and effective communication; 9. Have AI abide by these guidelines in order for it to become a quasi-member of society. Note, in particular, the Japanese twist of the last guideline.

¹³ See by way of example V Mnih, and others, 'Human-Level Control through Deep Reinforcement Learning' (2015) 518 *Nature* (26 February 2015) 529–533; see also B Schölkopf, 'Learning to See and Act' (2015) 518 *Nature* (26 February 2015) 486–487; and D Silver and others 'Mastering the Game of Go with Deep Neural Networks and Tree Search' (2016) 529 *Nature* (28 January 2016) 484–489. The Darpa Challenges also significantly pushed research forward, see T Burri, 'The Politics of Robot Autonomy' (2016) 7 *European Journal of Risk Regulation* (2) 341–360. In robotics, a certain amount of hysteria has been created by Boston Dynamics' videos. An early example is the video about the Atlas robot: Boston Dynamics, 'Atlas, the Next Generation' (YouTube, 23 February 2016) www.youtube.com/watch?v=rVlhMCQgDkY&app=desktop. But it is not all hype and hysteria, see already GA Pratt, 'Is a Cambrian Explosion Coming for Robotics?' (2015) 29 *Journal of Economic Perspectives* (3 (Summer 2015)) 51–60.

¹⁴ Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be deemed to be Excessively Injurious or to have Indiscriminate Effects (with Protocols I, II, and III), 1342 UNTS 163 (English), 10 October 1980.

¹⁵ This discussion was spurred on by a report: Human Rights Watch and Harvard International Human Rights Clinic, 'Losing Humanity: The Case against Killer Robots' (HRW, 19 November 2012) www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots, and an international civil society campaign, the Campaign to Stop Killer Robots (see www.stopkillerrobots.org), in which from the beginning researchers such as P Asaro, R Sparrow, N Sharkey, and others were involved; the International Committee for Robot Arms Control (ICRAC, see www.icrac.net) also campaigned against Killer Robots. Much of the influential legal work within the context of the Campaign goes back to B Docherty, e.g. the report just mentioned or B Docherty, 'Mind the Gap: The Lack of Accountability for Killer Robots' (HRW, 9 April 2015) www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots; B Docherty, 'Precedent for Preemption: The Ban on Blinding Lasers as a Model for a Killer Robots Prohibition' (HRW, 8 November 2015) www.hrw.org/news/2015/11/08/precedent-preemption-ban-blinding-lasers-model-killer-robots-prohibition. The issue of autonomous weapons systems had previously been addressed by Philip Alston: UNCHR, 'Interim Report by UN Special Rapporteur on extrajudicial, summary or arbitrary executions, Philip Alston' (2010) UN Doc A/65/321; see also P Alston, 'Lethal Robotic Technologies: The Implications for Human Rights and International Humanitarian Law' (2011) 21 *Journal of Law, Information and Science* 35–60; and later by Christof Heyns: UNCHR, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, (2013) UN Doc A/HRC/23/47; for scholarship, see A Leveringhaus, *Ethics and Autonomous Weapons* (2016).

¹⁶ The discussion of cyber warfare took a different path. See most recently, D Trusilo and T Burri, 'Ethical Artificial Intelligence: An Approach to Evaluating Disembodied Autonomous Systems' in R Liivoja and A Väljataga (eds), *Autonomous Cyber Capabilities under International Law* (2021) 51–66 (hereafter Trusilo and Burri, 'Ethical AI').

¹⁷ For a discussion of embodiment from a philosophical perspective, see C Durt, 'The Computation of Bodily, Embodied, and Virtual Reality' (2020) 1 *Phänomenologische Forschungen* 25–39 www.durt.de/publications/bodily-embodied-and-virtual-reality/.

¹⁸ Defence has meanwhile gone beyond autonomy to consider also AI. Contrast the early US Department of Defence, 'Directive on Autonomy in Weapon Systems' (DoD, 21 November 2012, amended 8 May 2017) www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf with the recent Defense Innovation Board, 'AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense' (DoD, 24 February 2020) 12 https://media.defense.gov/2019/Oct/31/2002204458/1/1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF: 'The important thing to consider going forward is that however DoD integrates AI into autonomous systems, whether or not they are weapons systems, sharp ethical and technical distinctions between AI and autonomy may begin to blur, and the Department should consider the interaction between AI and autonomy to ensure that legal and ethical dimensions are considered and addressed.' The Report addresses AI within the Department of Defense in general, not just in combat. It posits five key aspects which should inform the Department of Defense's engagement with AI: Responsible, equitable, traceable, reliable, governable. ('Equitable' refers to what is in other documents often

community became fixated on an exclusive and exotic aspect – namely physical (‘kinetic’) autonomous weapons systems – while the technological development was more comprehensive. Despite their narrow focus, the seven years of discussions in Geneva have yielded few concrete results, other than a great deal of publicity.¹⁹

At about the same time, autonomous cars also became the subject of ethical discussion. This discussion, however, soon got bogged down in largely theoretical, though fascinating, ethical dilemmas, such as the trolley problem.²⁰ However, unlike those gathered in Geneva to ponder autonomous weapons systems, those intent on putting autonomous cars on the road were pragmatic. They found ways of generating meaningful output that could be implemented.²¹

In 2017, the broader public beyond academic and professional circles became aware of the promises and perils of AI. Civil society began to discuss the ethics of AI and soon produced tangible output.²² Actionable principles were also proposed on behalf of

called ‘fairness’ or ‘avoidance of bias’, terms which, according to the report, may be misleading in defence, see p 31). See also HM Roff, ‘Artificial Intelligence: Power to the People’ (2019) 33 *Ethics and International Affairs* 127, 128–133, for a distinction between automation, autonomy, and AI.

¹⁹ The output consists of eleven high-level principles on autonomous weapons systems: Alliance for Multilateralism on Lethal Autonomous Weapons Systems (LAWS), ‘Eleven Guiding Principles on Lethal Autonomous Weapons Systems’ (*Alliance for Multilateralism*, 2020) <https://multilateralism.org/wp-content/uploads/2020/04/declaration-on-lethal-autonomous-weapons-systems-laws.pdf> (hereafter *Eleven Guiding Principles on Lethal Autonomous Weapons*); for the positions of states within CCW and the status quo of the discussions, see D Lewis, ‘An Enduring Impasse on Autonomous Weapons’ (*Just Security*, 28 September 2020) www.justsecurity.org/72610/an-enduring-impasse-on-autonomous-weapons/; for a thorough discussion of autonomous weapons systems and AI see AL Schuller, ‘At the Crossroads of Control: The Intersection of Artificial Intelligence and Autonomous Weapons Systems with International Humanitarian Law’ (2017) 8 *Harvard National Security Journal* (2) 379–425; see also SS Hua, ‘Machine Learning Weapons and International Humanitarian Law: Rethinking Meaningful Human Control’ (2019) 51 *Georgetown Journal of International Law* 117–146.

²⁰ See JF Bonnefon, A Shariff, and I Rahwan, ‘The Social Dilemma of Autonomous Vehicles’ (2016) 352 *Science* (6293) 1573–1576; E Awad and others, ‘The Moral Machine Experiment’ (2018) 563 *Nature* 59–64.

²¹ Note in particular, Ethics Commission of the Federal Ministry of Transport and Digital Infrastructure, ‘Automated and Connected Driving’ (BMVI, June 2017) www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile. This report pinpointed 20 detailed principles. The principles stated clearly that autonomous driving was ethically justified under certain conditions, even if the result of autonomous driving was that persons may occasionally be killed (see principles 2, 8, and 9). See also A von Ungem-Sternberg, ‘Autonomous Driving: Regulatory Challenges Raised by Artificial Decision-Making and Tragic Choices’ in W Barfield and U Pagallo (eds), *Research Handbook on the Law of Artificial Intelligence* (2017) 251–278.

²² The Future Society in Policy Research, The Law & Society Initiative, ‘Principles for the Governance of AI’ (*The Future Society*, 15 July 2017) <https://thefuturesociety.org/the-law-society-initiative/> (under ‘learn more’); University of Montreal, ‘Montreal Declaration for a Responsible Development of Artificial Intelligence’ (*Montréal Declaration Responsible AI*, 2018) https://docs.wixstatic.com/ugd/ebc3a3_c5c1c196fc164756afb92466c081d7ae.pdf (hereafter ‘Montreal Declaration for AI’) was one of the first documents to examine the societal implications of AI, putting forward a very broad and largely aspirational set of principles, the gist being: 1. Increase well-being (with 5 sub-principles); 2. Respect people’s autonomy and increase their control over lives (6 sub-principles); 3. Protect privacy and intimacy (8); 4. Maintain bonds of solidarity between people and generations (6); 5. Democratic participation in AI: it must be intelligible, justifiable, and accessible, while subject to democratic scrutiny, debate, and control (10); 6. Contribute to just and equitable society (7); 7. Maintain diversity, do not restrict choice and experience (6); 8. Prudence: exercise caution in development, anticipate adverse consequences (5); 9. Do not lessen human responsibility (5); 10. Ensure sustainability of planet (4). Compare with: Amnesty International and Access Now, ‘The Toronto Declaration: Protecting the Right to Equality and Non-Discrimination in Machine Learning Systems’ (16 May 2018) www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf (hereafter ‘Toronto Declaration’) which, although put together by non-governmental organizations, is more in the nature of an academic legal text and not easily summarized. It emphasizes the duties of states to identify risks, ensure transparency and accountability, enforce oversight, promote equality, and hold the private sector to account. Similar duties are incumbent on private actors, though they are less firm. The right to effective remedy is also emphasized. Compare

women²³ and labour,²⁴ and *AlgorithmWatch*, a now notable non-governmental organization, was founded.²⁵

In step with civil society, private companies adopted ethical principles concerning AI.²⁶ Such principles took different shapes depending on companies' fields of business. The principles

also with The Public Voice, 'Universal Guidelines for Artificial Intelligence' (*The Public Voice*, 23 October 2018) <https://thepublicvoice.org/ai-universal-guidelines/>.

²³ Women Leading in AI, '10 Principles of Responsible AI' (*Women Leading in AI*, 2019) <https://womenleadinginai.org/wp-content/uploads/2019/02/WLiAI-Report-2019.pdf>. This initiative did not look at AI strictly from a gender, but a broader societal perspective. The 10 principles can be summarized as follows: 1. Mirror the regulatory approach for the pharmaceutical sector; 2. Establish an AI regulatory body with powers inter alia to: audit algorithms, investigate complaints, issue fines for breaches of the General Data Protection Regulation, the law and equality, and ensure algorithms are explainable; 3. Introduce 'Certificate of Fairness for AI systems'; 4. Require 'Algorithm Impact Assessment' when AI is employed with impact on individuals; 5. In public sector, inform when decisions are made by machines; 6. Reduce liability when 'Certificate of Fairness' is given; 7. Compel companies to bring their workforce with them; 8. Establish digital skills funds to be fed by companies; 9. Carry out skills audit to identify relevant skills for transition; 10. Establish education and training programme, especially to encourage women and underrepresented sections of society.

²⁴ UNI Global Union, '10 Principles for Ethical AI, UNI Global Union Future World of Work' (*The Future World of Work*, 2017) www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf; summarized: 1. Transparency; 2. Equip with black box; 3. Serve people/planet; 4. Humans must be in command, incl. responsibility, safety, compliance with privacy and law; 5. Avoid bias in AI; 6. Share benefits; 7. Just transition for workforce and support for human rights; 8. Establish global multi-stakeholder governance mechanism for work and AI; 9. Ban responsibility of robots; 10. Ban autonomous weapons.

²⁵ See <https://algorithmwatch.org/en/transparency/>; AlgorithmWatch provides a useful database bringing together ethical guidelines on AI: <https://inventory.algorithmwatch.org/>. In 2017, the AI Now Institute at New York University, which conducts research on societal aspects of AI, was also established (see www.ainowinstitute.org). Various 'research agendas' have by now been published: J Whittlestone and others, 'Ethical and Societal Implications of Algorithms, Data, and Artificial Intelligence: A Roadmap for Research' (*Nuffield Foundation*, 2019) www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf (with a useful literature review in appendix 1 and a review of select ethics principles in appendix 2); A Dafoe, 'AI Governance: A Research Agenda' (*Future of Humanity Institute*, 2018) www.fhi.ox.ac.uk/wp-content/uploads/GovAI/Agenda.pdf, which broadly focuses on economics and political science research. Compare with OpenAI which is on a 'mission' to ensure that general AI will be beneficial. For this purpose, it conducts research on AI based on its own ethical Charter: OpenAI, OpenAI Charter (*Open AI*, 9 April 2018) <https://openai.com/charter/> (hereafter OpenAI Charter); in brief, the principles of the Charter are: Ensure general AI benefits all, avoid uses that harm or concentrate power; primary duty to humanity, minimize conflicts of interest that compromise broad benefit; do the research that makes general AI safe; if late-stage development of general AI becomes a competitive race without time for precaution, stop competing and assist the other project; leadership in technology, policy, and safety advocacy is not enough; AI will impact before general AI, so lead there too; cooperate actively, create global community; provide public goods that help society navigate towards general AI; for now, publish most AI research, but later probably not for safety reasons.

²⁶ See Intel, 'AI Public Policy Opportunity' (*Intel*, 2017) <https://blogs.intel.com/policy/files/2017/10/Intel-Artificial-Intelligence-Public-Policy-White-Paper-2017.pdf> summed up: 1. Foster innovation and open development; 2. Create new human employment and protect people's welfare; 3. Liberate data responsibly; 4. Rethink privacy; 5. Require accountability for ethical design and implementation. Further examples include Sage, 'The Ethics of Code: Developing AI for Business with Five Core Principles' (*Sage*, 2017) www.sage.com/~media/group/files/business-builders/business-builders-ethics-of-code.pdf?la=en&hash=CB4DF0EB6CCB15F55E72EBB3CD5D526B (hereafter Sage, 'The Ethics of Code'), in brief: 1. Reflect diversity, avoid bias; 2. Accountable AI, but also accountable users; AI must not be too clever to be held accountable; 3. Reward AI for aligning with human values through reinforcement learning; 4. AI should level playing field: democratize access, especially for disabled persons; 5. AI replaces, but must also create work: humans should focus on what they are good at; Google, 'Artificial Intelligence at Google: Our Principles' (*Google*, 2018) <https://ai.google/principles/> (hereafter Google, 'AI Principles'); in brief: 1. Be socially beneficial and thoughtfully evaluate when to make technology available on non-commercial basis; 2. Avoid bias; 3. Build and test for safety; 4. Be accountable to people, i.e. offer feedback, explanation, and appeal; subject AI to human direction and control; 5. Incorporate privacy design principles; 6. Uphold high standard of scientific excellence; 7. Use of AI must accord with these principles; 8. No-go areas: technology likely to cause overall harm; weapons; technology for surveillance violating internationally accepted norms; technology whose purpose violates international law and human rights – though this 'point 8' may evolve; IBM, 'Everyday Ethics for Artificial Intelligence' (*IBM*, 2018) www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf (hereafter IBM, 'Ethics for AI'); in brief: 1. Be accountable, i.e.

embody a certain degree of self-commitment, which is not subject to outside verification, though.²⁷ Parts of the private sector and the third sector have also joined forces, most prominently in the Partnership on AI and its tenets on AI.²⁸

The development has not come to a halt today. Various organizations continue to mull over ethical norms to govern AI.²⁹ However, most early proponents of such norms have moved from the formation stage to the implementation stage. Private companies are currently applying the principles to which they unilaterally subscribed. After having issued one of the first documents on ethical norms,³⁰ the IEEE is now developing concrete technical standards to be applied by developers to specific applications of AI.³¹ ISO, another professional organization, is currently setting such standards as well.³² Domestic courts and authorities are adjudicating the first cases on AI.³³

At this point, it is worth pausing for a moment. The current section sketched a process in which multiple actors shaped and formed ethical norms on AI and are now implementing them. (As Section IV will explain, states have not been absent from this process.) This section could now go on to distil the essence of the ethical norms. This would make sense as the ethics remain

understand accountability, keep records, understand the law. 2. Align with user values, inter alia by bringing in policy makers and academics; 3. Keep it explainable, i.e., allow for user questions and make AI reviewable; 4. Minimize bias and promote inclusion. 5. Protect users' data rights, adhere to national and international rights laws.

²⁷ AI Now, 'AI Now 2017 Report' (AI Now Institute, 2017) https://ainowinstitute.org/AI_Now_2017_Report.pdf, recommendation no 10: 'Ethical codes [...] should be accompanied by strong oversight and accountability mechanisms.' (p 2); see also AI Now, 'AI Now 2018 Report' (AI Now Institute, 2018) https://ainowinstitute.org/AI_Now_2018_Report.pdf, recommendation no 3: 'The AI industry urgently needs new approaches to governance.' (p 4).

²⁸ Partnership on AI, 'Tenets' www.partnershiponai.org/tenets/ (hereafter Partnership on AI, 'Tenets'), in summary: 1. Benefit and empower as many people as possible; 2. Educate and listen, inform; 3. Be committed to open research and dialogue on the ethical, social, economic, and legal implications of AI; 4. Research and development need to be actively engaged with, and accountable to, stakeholders; 5. Engage with, and have representation of, stakeholders in the business community; 6. Maximize benefits and address challenges by: protecting privacy and security; understanding and respecting interests of all parties impacted; ensuring that the AI community remains socially responsible, sensitive and engaged; ensuring that AI is robust, reliable, trustworthy, and secure; opposing AI that would violate international conventions and human rights; and promoting safeguards and technology that do no harm; 7. Be understandable and interpretable for people for purposes of explaining the technology; 8. Strive for a culture of cooperation, trust, and openness among AI scientists and engineers.

²⁹ See, for instance, Pontifical Academy for Life, Microsoft, IBM, FAO and Ministry of Innovation (Italian Government), 'Rome Call for AI Ethics' (Rome Call, 28 February 2020) www.romecall.org.

³⁰ IEEE, 'Ethically Aligned Design' (n 10).

³¹ See the IEEE P7000 standards series, e.g. IEEE SA, IEEE P7000 - Draft Model Process for Addressing Ethical Concerns During System Design (IEEE, 30 June 2016) <https://standards.ieee.org/project/7000.html>; The IEEE considers standard setting with regard to AI unprecedented: 'This is the first series of standards in the history of the IEEE Standards Association that explicitly focuses on societal and ethical issues associated with a certain field of technology'; IEEE, 'Ethically Aligned Design' (n 10) 283; for the type of standard that is necessary, see D Danks, AJ London, 'Regulating Autonomous Systems: Beyond Standards', (2017) 32 *IEEE Intelligent Systems* 88.

³² See ISO, 'Standards by ISO / IEC JTC 1 / SC 42. Artificial Intelligence' www.iso.org/committee/6794475/x/catalogue/p/o/u/h/w/o/d/o.

³³ See UK High Court, R (Bridges) v CCSWP and SSHD [2019] EWHC 2341 (Admin); UK High Court, R (Bridges) v CCSWP and SSHD [2020] EWCA Civ 1058; Tribunal Administratif de Marseille, La Quadrature du Net, No. 1901249 (27 Nov. 2020); Swedish Data Protection Authority, 'Supervision pursuant to the General Data Protection Regulation (EU) 2016/679 – facial recognition used to monitor attendance of students' (DI-2019-2221, 20 August 2019) <imy.se/globalassets/dokument/beslut/facial-recognition-used-to-monitor-the-attendance-of-students.pdf>; a number of non-governmental organisations are bringing an action against Clearview AI Inc., which sells facial recognition software, for violation of data protection law, see <https://privacyinternational.org/legal-action/challenge-against-clearview-ai-europe>. A global inventory listing incidents involving AI that have taken place so far includes more than 600 entries to date: AIAIC repository: https://docs.google.com/spreadsheets/d/1Bn55B4xz21-RgdR8BBb2ton_4rzLGxFADMVW0PYI/edit#gid=888071280; compare with AI Incident Database, 'All Incident Reports' (7 June 2021) <https://incidentdatabase.ai/>, which is run by the Partnership on AI and includes 100 incidents.

unconsolidated and fuzzy. But much important work has already been done in this direction.³⁴ In fact, for present purposes, no further efforts are necessary because, while norms remain vague, they have now begun to merge into domestic law. However, the diffusion of ethical norms is far from being a linear and straightforward process with clear causes. Instead, it is multidirectional, multivariate, gradual, and open-ended, with plenty of back and forth. Hence, the next section, as it looks at norm diffusion from the incoming end, in other words, from the perspectives of states and domestic law, is best read as a continuation of the present section. The developments outlined have also occurred in parallel to those in municipal law, which are the topic of the next section.

III. DIFFUSION OF ETHICAL NORMS INTO DOMESTIC LAW: THE NEW REGULATION OF THE EUROPEAN UNION ON AI

A relevant sign of diffusion into domestic law is states' first engagement with ethics and AI. For some states, including China, France, Germany, and the United States, such engagement began relatively early with the adoption of AI strategies³⁵ in which ethical norms figured more or less prominently. The French president, for instance, stated a commitment to establish an ethics framework.³⁶ China, in its strategy, formulated the aim to '[d]evelop laws, regulations, and ethical norms that promote the development of AI'.³⁷ Germany's strategy was to task a commission to come up with recommendations concerning ethics.³⁸ The US strategy, meanwhile, was largely silent on ethics.³⁹

³⁴ A Jobin, M Ienca, and E Vayena, 'The Global Landscape of AI Ethics Guidelines' (2019) 1 *Nature Machine Intelligence* (2019) 389–399; J Fjeld and others, 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI' (*Berkman Klein Center for Internet & Society*, 2020) <http://nrs.harvard.edu/urn-3:HUL.InstRepos:42160420>.

³⁵ State Council of the People's Republic of China, 'A Next Generation Artificial Intelligence Development Plan' (*New America*, 20 July 2017) www.newamerica.org/documents/1959/translation-fulltext-8.1.17.pdf (hereafter China, 'AI Development Plan'); President of the French Republic, 'The President of the French Republic Presented His Vision and Strategy to Make France a Leader in AI at the Collège de France on 29 March 2018' (*AI for Humanity*, 2018) www.aiforhumanity.fr/en/ (hereafter French Republic, 'Strategy to Make France a Leader in AI'); Federal Government of Germany, 'Artificial Intelligence Strategy' (*The Federal Government*, November 2018) www.ki-strategie-deutschland.de/home.html?file=files/downloads/Nationale_KI-Strategie_engl.pdf (hereafter Germany, 'AI Strategy'); US President, 'Executive Order on Maintaining American Leadership in Artificial Intelligence' (2019) E.O. 13859 of Feb 11, 2019, 84 FR 3967 (hereafter US President, 'Executive Order on Leadership in AI'). According to T Dutton, 'An Overview of National AI Strategies' (*Medium*, 28 June 2018) <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd> which contains a useful list of national AI strategies, Canada was the first state to put forward such a national strategy in the year 2017. Yet it remains unclear what exactly constitutes a 'strategy'. In any case, the documents published by the Obama Administration in 2016 (see n 38) already contained many elements of a 'strategy'.

³⁶ French Republic, 'Strategy to Make France a Leader in AI' (n 35) third commitment.

³⁷ China, 'AI Development Plan' (n 35) Section V 1; the text accompanying this aim is more concrete. It recommends addressing traceability and accountability; to launch research on AI behaviour science and ethics; and 'establish an ethical and moral multi-level judgment structure and human-computer collaboration ethical framework'. China is also committed to 'actively participate in global governance of AI, strengthen the study of major international common problems such as robot alienation and safety supervision, deepen international cooperation on AI laws and regulations, international rules and so on, and jointly cope with global challenges'.

³⁸ Germany, 'AI Strategy' (n 35) 4, 37, 38. The data ethics commission ('Datenethikkommission') in response published its report in October 2019: Datenethikkommission, 'Gutachten' (*BMI*, October 2019) www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf?__blob=publicationFile&v=4. The report deals comprehensively on 240 pages with 'digitization', not just AI, and includes 75 recommendations to move forward. An economic assessment of the proposals in the report would be necessary though. The report seems quite 'big' on regulation.

³⁹ The US strategy merely stated as one of five guiding principles: 'The United States must foster public trust and confidence in AI technologies and protect civil liberties, privacy, and American values in their application in order to fully realize the potential of AI technologies for the American people.' (US President, 'Executive Order on Leadership in AI' (n 35) section 1(d); compare with National Science and Technology Council, 'Preparing for the Future of

Some state legislative organs also addressed the ethics of AI early on, most notably, the comprehensive report published by the United Kingdom House of Lords in 2018.⁴⁰ It, among other things, recommended elaborating an AI code to provide ethical guidance and a ‘basis for statutory regulation, if and when this is determined to be necessary’.⁴¹ The UK report also suggested five ethical principles as a basis for further work.⁴² In a similar vein, the Villani report, which had preceded the French presidential strategy, identified five ethical imperatives.⁴³

In the EU, a report drafted within the European Parliament in 2016 drew attention to the need to examine ethics further.⁴⁴ It dealt with robotics because AI was not yet a priority and included a code of rudimentary ethical principles to be observed by researchers. In 2017, the European Parliament adopted the report as a resolution,⁴⁵ putting pressure on the Commission to propose legislation.⁴⁶ In 2018, the Commission published a strategy on AI with a threefold

Artificial Intelligence’ (*The White House, President Barack Obama*, October 2016) https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf which had been published before and addressed transparency, fairness, and efficacy of systems in recommendations nos 16 and 17 and ethics in education curricula in recommendation no 20, and National Science and Technology Council, ‘The National Artificial Intelligence Research and Development Strategic Plan’, (*The White House, President Barack Obama*, October 2016) https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf, which was published on the same day as *Preparing for the Future of Artificial Intelligence*, p. 3: ‘understand and address the ethical, legal, and societal implications of AI’ is a research priority according to strategy no. 3. See also the webpage of the US government on AI which has recently gone live: www.ai.gov/.

⁴⁰ House of Lords (Select Committee on Artificial Intelligence), ‘AI in the UK: Ready, Willing and Able?’ (*UK Parliament*, 16 April 2018) <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf> (hereafter House of Lords, ‘AI in the UK’).

⁴¹ House of Lords, ‘AI in the UK’ (n 40) para 420.

⁴² House of Lords, ‘AI in the UK’ (n 40) para 417, in brief: 1. Development of AI for common good and humanity; 2. Intelligibility and fairness; 3. Use of AI should not diminish data rights or privacy; 4. Individuals’ right to be educated to flourish mentally, emotionally and economically alongside AI; 5. The autonomous power to hurt, destroy, or deceive human beings should never be vested in AI. In the United Kingdom, further work also addressed the use of facial recognition technology: Biometrics and Forensics Ethics Group (BFEG UK government), ‘Interim Report of BFEG Facial Recognition Working Group’ (OGL, February 2019) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/781745/Facial_Recognition_Briefing_BFEG_February_2019.pdf.

According to this report, facial recognition: 1. Is only permissible when in public interest; 2. Justifiable only if effective; 3. Should not involve or exhibit bias; 4. Should be deployed in even-handed ways: for example, not target certain events only (impartiality); 5. Should be a last resort: No other less invasive alternative, minimizing interference with lawful behaviour (necessity). Also, 6. Benefits must be proportionate to loss of liberty and privacy; 7. Humans must be impartial, accountable, oversights, esp. when constructing watch lists; and 8. Public consultation and rationale are necessary for trust. Finally, 9. Could resources be used better elsewhere?

⁴³ C Villani, ‘For a Meaningful Artificial Intelligence – Towards a French and European Strategy’ (*AI for Humanity*, March 2018) www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf 113–114; in summary: 1. transparency and auditability; 2. Rights and freedoms need to be adapted in order to forestall potential abuse; 3. Responsibility; 4. Creation of a diverse and inclusive social forum for discussion; 5. Politicization of the issues linked to technology. Compare with D Dawson and others, ‘Artificial Intelligence – Australia’s Ethics Framework, A Discussion Paper’ (2019) https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/supporting_documents/ArtificialIntelligenceethicsframeworkdiscussionpaper.pdf 6, which, in a nutshell, proposed the following ethics guidelines: 1. Generate net benefits; 2. Civilian systems should do no harm; 3. Regulatory and legal compliance; 4. Protection of privacy; 5. Fairness: no unfair discrimination, particular attention to be given to training data; 6. Transparency and explainability; 7. Contestability; 8. Accountability, even if harm was unintended.

⁴⁴ Draft Report with recommendations to the Commission on Civil Law Rules on Robotics, 2015/2103(INL), 23 May 2016; the report was marked by an alarmist undertone.

⁴⁵ Resolution with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), European Parliament, P8_TA (2017)0051, 16 February 2018.

⁴⁶ *Ibid.*, para 65.

aim, one of which was to ensure ‘an appropriate legal and ethical framework’.⁴⁷ The Commission consequently mandated a group of experts who suggested guidelines for ‘trustworthy’ AI one year later.⁴⁸ These guidelines explicitly drew on work previously done within the institutions.⁴⁹ The guidelines refrained from interfering with the *lex lata*,⁵⁰ including the General Data Protection Regulation⁵¹.

In 2019, following the guidelines for trustworthy AI, the Commission published a White Paper on AI⁵², laying the foundation for the legislative proposal to be tabled a year later. The White Paper, which attracted much attention,⁵³ recommended a horizontal approach to AI with general principles included in a single legislative act applicable to any kind of AI, thus rejecting the alternative of adapting existing (or adopting several new) sectorial acts. The White Paper

⁴⁷ 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, Artificial Intelligence for Europe, European Commission, 25 April 2018, section 1 toward the end.

⁴⁸ High-Level Expert Group on Artificial Intelligence, ‘Ethics Guidelines for Trustworthy AI’ (8 April 2019) www.ai.bsa.org/wp-content/uploads/2019/09/AIHLEG_EthicsGuidelinesforTrustworthyAI-EN.pdf (hereafter: ‘Ethics Guidelines for Trustworthy AI’). The Guidelines distinguish between foundations of trustworthy AI which include four ethical principles, namely 1. Respect for human autonomy, 2. Prevention of harm, 3. Fairness, 4. Explicability (12 *et seq*) and seven requirements for their realization, namely 1. Human agency and oversight, 2. Technical robustness and safety, 3. Privacy and data governance, 4. Transparency, 5. Diversity, non-discrimination, fairness, 6. Societal and environmental well-being and 7. Accountability.

⁴⁹ Notably European Group on Ethics in Science and New Technologies (EGE), ‘Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems’ (9 March 2018) <https://op.europa.eu/en/publication-detail/-/publication/dfcbe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-e-78120382>>. Another initiative within the wider sphere of the EU worked in parallel with the Commission’s High-Level Expert Group and published a set of principles: L Floridi and others, ‘AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations’ (2018) 28 *Minds and Machines* 689.

⁵⁰ See Ethics Guidelines for Trustworthy AI (n 48) 6: ‘The Guidelines do not explicitly deal with the first component of Trustworthy AI (lawful AI), but instead aim to offer guidance on fostering and securing the second and third components (ethical and robust AI).’ And 10: ‘Understood as legally enforceable rights, fundamental rights therefore fall under the first component of Trustworthy AI (lawful AI), which safeguards compliance with the law. Understood as the rights of everyone, rooted in the inherent moral status of human beings, they also underpin the second component of Trustworthy AI (ethical AI), dealing with ethical norms that are not necessarily legally binding yet crucial to ensure trustworthiness.’

⁵¹ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1 (GDPR). The General Data Protection Regulation, in Article 22 regulates automated decision making and therefore one aspect of AI; however, the effectiveness of the Article is limited by the scope of Regulation as well as loopholes in paragraph 2. Article 22 is entitled ‘Automated Individual Decision-Making, Including Profiling’ and reads as follows: ‘1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her. 2. Paragraph 1 shall not apply if the decision: (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller; (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests; or (c) is based on the data subject’s explicit consent. 3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision; 4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests are in place.’ For an international legal perspective on the General Data Protection Regulation, see the Symposium on: ‘The GDPR in International Law’ (6 January 2020) *AJIL Unbound* 114.

⁵² European Commission, *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*, European Commission (White Paper, COM(2020) 65 final, 2020) (hereafter White Paper on AI).

⁵³ The public consultation on the White Paper on AI (n 52) attracted a wide range of comments, see e.g. Google, ‘Consultation on the White Paper on AI – a European Approach’ (Google, 28 May 2020) www.blog.google/docs/consultations/77/Googles_submission_to_EC_AI_consultation_1.pdf.

suggested regulating AI based on risk: the higher the risk of an AI application, the more regulation was necessary.⁵⁴

On 21 April 2021, based on the White Paper, the Commission presented a Proposal for a regulation on AI⁵⁵. The Commission's Proposal marks a crucial moment, for it represents the first formal step – globally, it seems – in a process that will ultimately lead to binding domestic legislation on AI. It is a sign of the absorption of ethical norms on AI by domestic law – in other words, of norm diffusion. While the risk-based regulatory approach adopted from the White Paper was by and large absent in the ethics documents discussed in the previous section, many of the substantive obligations in the proposed regulation reflect the same ethical norms.

The Commission proposed distinguishing three categories of AI, namely: certain 'practices' of AI that the proposed regulation prohibits; high-risk AI, which it regulates in-depth; and low-risk AI required to be flagged.⁵⁶ While the prohibition against using AI in specific ways (banned 'practices')⁵⁷ attracts much attention, practically, the regulation of high-risk AI will be more relevant. Annexes II and III to the proposed regulation determine whether an AI qualifies as high-risk.⁵⁸ The proposed regulation imposes a series of duties on those who place such high-risk AI on the market.⁵⁹

The regulatory focus on risky AI has the consequence, on the flip side, that not all AI is subject to the same degree of regulation. Indeed, the vast majority of AI is subject merely to the duty to ensure some degree of transparency. However, an AI that now appears to qualify as low-risk

⁵⁴ White Paper on AI (n 52) 17: an application of AI should be considered high-risk, when it is situated in a sensitive domain, e.g. health care, and presents a concrete risk.

⁵⁵ Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, European Commission, COM (2021) 206 final, 21 April 2021, in the following: the Proposal or the proposed regulation.

⁵⁶ See Article 52 of the proposed regulation which states a relatively light transparency obligation with regard to AI not presenting high risks ('certain AI systems', according to Article 52).

⁵⁷ The regulation proposes to ban the practice of AI: a) to materially distort a person's behaviour (a draft leaked earlier had called this 'manipulation'); b) to exploit the vulnerabilities of a specific group of persons ('targeting' of vulnerable groups, according to the leaked draft); c) social scoring by the public authorities, and d) for live remote biometric identification in public places (see article 5(1)(a)–(d) of the proposed regulation). The regulation does not preclude the development of AI, even if it could eventually be used in ways the regulation prohibits. A pathway is required in the case of letters a and b: the practices are only prohibited if they are at least likely to cause a person physical or psychological harm. The ban of biometric identification according to letter d is subject to a public security exception pursuant to Article 5(2).

⁵⁸ The definition of AI in annex I appears to be in accordance with how the term is understood in the computer sciences (compare S Russell and P Norvig, *Artificial Intelligence: A Modern Approach* (3rd ed., 2014), but it is a broad definition that lawyers may read differently than computer scientists and the elements added in Article 3(1) of the proposed regulation distort it to some degree. Annex II lists legislative acts of the Union; if an act listed applies (e.g., in case of medical devices or toys), any AI used in this context is to be considered high-risk. Annex III relies on domains in conjunction with concrete, intended uses. It lists the following domains: remote biometric identification systems (if not banned by article 5), critical infrastructure, educational institutions, employment, essential public and private services, law enforcement and criminal law, management of migration, asylum, and border control, as well as assistance of judicial authorities. Specific uses listed under these domains are caught as high-risk AI. For instance, AI is considered high-risk when it is intended to be used for predictive policing (use) in law enforcement (domain). The Commission, jointly with the Parliament and the Council, is delegated the power to add further uses within the existing domains, which, in turn, could only be added to by means of a full legislative amendment; the Commission's power is subject to an assessment of potential harm (see Articles 7 and 73 of the proposed regulation).

⁵⁹ Mostly the 'provider' will be the person who puts an AI on the market, according to Article 16 of the proposed regulation; sometimes it is the importer, the distributor or another third party, according to Articles 26–28; Article 3(2) defines a provider as 'a natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge'.

under the proposed regulation could become high-risk after a minor change in use intention. Hence, given the versatility of AI, the duties applicable to high-risk AI have to be factored in even in the development of AI in low-risk domains. One example is an image recognition algorithm that *per se* qualifies as low-risk under the regulation. However, if it were later used for facial recognition, the more onerous duties concerning high-risk AI would become applicable. Such development must be anticipated at an early stage to ensure compliance with the regulation throughout the life cycle of AI. Hence, regulatory spill-over from high-risk into low-risk domains of AI is likely. Consequently, the proposed regulation exerts a broader compliance pull than one might expect at first glance, given the specific, narrow focus of the regulation on high-risk AI.

Categorization aside, the substantive duties imposed on those who put high-risk AI on the market are most interesting from the perspective of ethical norm diffusion. The proposed regulation includes four bundles of obligations.

The first bundle concerns *data* and is laid down in Article 10 of the proposed regulation. When AI is trained with data (though not only then⁶⁰), Article 10 of the proposed regulation requires ‘appropriate data governance and management practices’, in particular concerning design choices; data collection; data preparation; assumptions concerning that which data measures and represents; assessment of availability, quantity, and suitability of data; ‘examination in view of possible bias’; and identification of gaps and shortcomings. In addition, the data itself must be relevant, representative, free of errors, and complete. It must also have ‘appropriate statistical properties’ regarding the persons on whom the AI is used. And it must take into account the ‘geographical, behavioural or functional setting’ in which the AI will be used.

The duties laid down in Article 10 on data mirror existing ethical norms, notably the imperative to avoid bias. The IEEE’s Charter discussed the issue of data bias.⁶¹ In an early set of principles addressed to professionals, avoidance of bias featured prominently; it also recommended keeping a description of data provenance.⁶² The Montreal Declaration recommended avoiding discrimination,⁶³ while the Toronto Declaration on human rights and machine learning had bias and discrimination squarely in view.⁶⁴ Likewise, some of the ethical norms the private sector had adopted addressed bias.⁶⁵ However, the ethical norms discussed in Section II generally refrained from addressing data and its governance as

⁶⁰ Article 10(6) of the proposed regulation transposes some of the requirements applicable to trained AI to AI that has not been trained.

⁶¹ IEEE, ‘Ethically Aligned Design’ (n 10) 188, recommending careful assessment of bias and integration of potentially disadvantaged groups in the process; Future of Life Institute, ‘Asilomar AI Principles’ (n 11) did not yet address bias explicitly.

⁶² USACM, ‘Algorithmic Transparency’ (n 12), principle no 1: ‘1. Awareness: Owners, designers, builders, users, and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society.’ Principle no 5 addressed ‘data provenance’. Compare Japanese Society for AI, ‘Guidelines’ (n 12) principle no 5 with a slightly broader scope.

⁶³ Montreal Declaration for AI (n 22) principle no 6.1: ‘AIS must be designed and trained so as not to create, reinforce, or reproduce discrimination based on – among other things – social, sexual, ethnic, cultural, or religious differences.’ See also principle no 7 concerning diversity; there are some data governance requirements in principle no 8 on prudence.

⁶⁴ Toronto Declaration (n 22) for instance, no 16. Not all documents laying down ethics principles discuss bias; OpenAI Charter (n 25) for instance, leaves bias aside and focuses on the safety of general AI.

⁶⁵ By way of example, Sage, ‘The Ethics of Code’ (n 26) principle no 1; Google, ‘AI Principles’ (n 26) principle no 2; IBM, ‘Ethics for AI’ (n 26) discusses fairness, including avoidance of bias, as one of five ethics principles (34–35); it also includes recommendations on how to handle data: ‘Your AI may be susceptible to different types of bias based on the type of data it ingests. Monitor training and results in order to quickly respond to issues. Test early and often.’ Partnership on AI, Tenets (n 28) on the other hand, only generically refers to human rights (see tenet no 6.e).

comprehensively as Article 10 of the proposed regulation. Instead, the ethical norms directly focused on avoidance of bias and discrimination.

The second bundle of obligations concerns *transparency* and is contained in Article 13 of the proposed regulation. The critical duty of Article 13 requires providers to ‘enable users to interpret [the] output’ of high-risk AI and ‘use it appropriately’⁶⁶. The article further stipulates that providers have to furnish information that is ‘concise, complete, correct and clear’⁶⁷, in particular regarding the ‘characteristics, capabilities and limitations of performance’ of a high-risk AI system.⁶⁸ These duties specifically relate to any known or foreseeable circumstance, including foreseeable misuse, which ‘may lead to risks to health and safety or fundamental rights’, and to performance on persons.⁶⁹

Transparency is an equally important desideratum of ethical norms, though it is sometimes addressed in terms of explainability or explicability. The IEEE’s Charter⁷⁰ and the Asilomar principles⁷¹ emphasized transparency to different degrees. Other guidelines encourage the production of explanations⁷² or appropriate and sufficient information,⁷³ or call for extensive transparency, justifiability, and intelligibility.⁷⁴ These references make it evident that ethical norms, though they are heterogeneous and vague, are in the process of being absorbed by EU law (norm diffusion).

The third bundle of obligations is contained in Article 15 of the proposed regulation. It requires high-risk AI to have an ‘appropriate level’ of *accuracy*, *robustness*, and *cybersecurity*.⁷⁵ Article 15 refrains from adding much detail but states that the AI must be resilient to deleterious environmental influences or nefarious third parties’ attempts to game it.⁷⁶

As with the first and second bundles, the aspects of high-risk AI addressed by Article 15 can be traced back to various ethical norms. The high-level principles of effectiveness and awareness of misuse in the IEEE’s Charter covered similar aspects.⁷⁷ The Asilomar principles addressed ‘safety’, but in a rather generic fashion.⁷⁸ Other principles emphasized both the need for safety in all things related to AI and the importance of preventing misuse.⁷⁹ Others focused on prudence, which more or less includes the aspects covered by Article 15.⁸⁰ Parts of the private sector also committed themselves to safe AI.⁸¹

⁶⁶ Article 13(1) of the proposed regulation.

⁶⁷ Article 13(2) of the proposed regulation.

⁶⁸ Article 13(3b) of the proposed regulation.

⁶⁹ Article 13(3b)(iii and iv) of the proposed regulation.

⁷⁰ IEEE, ‘Ethically Aligned Design’ (n 10) 11; transparency implies that the basis of a decision of an AI should ‘always be discoverable’.

⁷¹ Asilomar AI Principles (n 11) according to principle no 7, it must be possible to ascertain why an AI caused harm; according to principle no 8, any involvement in judicial decision making should be explainable and auditable.

⁷² USACM, ‘Algorithmic Transparency’ (n 12) principle no 4.

⁷³ Japanese Society for AI, ‘Guidelines’ (n 12) principle no 5 (addressing security).

⁷⁴ Montreal Declaration for AI (n 22) principle no 5, with 10 sub-principles addressing various aspects of transparency. See also The Toronto Declaration (n 22) which includes strong transparency obligations for states (para 32) and weaker obligations for the private sector (para 51).

⁷⁵ Article 15(1) of the proposed regulation.

⁷⁶ Article 15(3 and 4) of the proposed regulation.

⁷⁷ IEEE, ‘Ethically Aligned Design’ (n 10) 11, principles nos 4 and 7.

⁷⁸ Asilomar AI Principles (n 11) principle no 6.

⁷⁹ Japanese Society for AI, ‘Guidelines’ (n 12) principles nos 5 and 7.

⁸⁰ ‘Montreal Declaration for AI (n 22) principle no 8; The Toronto Declaration (n 22) has a strong focus on non-discrimination and human rights; it does not address the topics covered by Article 15 of the proposed regulation directly. Open AI Charter (n 25) stated a commitment to undertake the research to make AI safe in the long term.

⁸¹ E.g. Google, ‘AI Principles’ (n 26) principle no 3: ‘Be built and tested for safety’; IBM, ‘Ethics for AI’ (n 26) 42–45, addressed certain aspects of safety and misuse under ‘user data rights’. See also Partnership on AI, ‘Tenets’ (n 28) tenet

The fourth bundle contains obligations of a *procedural or managerial nature*. The proposed regulation places confidence in procedure to cope with the high risks of AI. The trust in procedure goes so far that substantive issues are addressed procedurally only. One such example is one of the cardinal obligations of the proposed regulation, namely the duty to manage risks according to Article 9. Article 9 obliges providers to maintain a comprehensive risk management system throughout the life cycle of high-risk AI. It aims at reducing the risks posed by the AI so that the risks are ‘judged acceptable’, even under conditions of foreseeable misuse.⁸² The means to reduce the risks are design, development, testing, mitigation and control measures, and provision of information to users. Instead of indicating which risks are to be ‘judged acceptable’, Article 9 trusts that risk reduction will result from a series of diligently executed, proper steps. However, procedural rules are not substantive rules. In and of themselves, they do not contain substantive guidance. In essence, Article 9 entrusts providers with the central ‘judgment’ of what is ‘acceptable’. Providers are granted liberty, while their obligations seem less onerous. At the same time, this liberty imposes a burden on them in that courts might not always validate their ‘judgment’ of what was ‘acceptable’ after harm has occurred. Would, for instance, private claims brought against the provider of an enormously beneficial AI be rejected after exceptionally high risks, which the provider managed and judged acceptable, have materialized?

Trust in procedure is also a mainstay of other provisions of the proposed regulation. An assessment of conformity with the proposed regulation has to be undertaken, but, here again, providers carry it out themselves in all but a few cases.⁸³ Providers have to register high-risk AI in a new EU-wide database.⁸⁴ Technical documentation and logs must be kept.⁸⁵ Human oversight is required – a notion that has a procedural connotation.⁸⁶ The regulation does not require substantive ‘human control’ as discussed within CCW for autonomous weapons systems.⁸⁷ Discrimination is not directly prohibited, but procedural transparency is supposed to contribute to preventing bias.⁸⁸ Such transparency may render high-risk AI interpretable, but a substantive right to explicable AI is missing.⁸⁹

The procedural and managerial obligations in the fourth bundle cannot easily be traced back to ethical norms. This is because of their procedural nature. Ethical norms are, in essence, substantive norms. Procedural obligations are geared towards implementation, yet implementation is not the standard domain of ethics (except for applied ethics which is yet to reach AI⁹⁰). Hence, while certain aspects of the fourth bundle mirror ethical norms, for example, the requirement to keep logs,⁹¹ none of them has called for a comprehensive risk management system.

no 6.d: ‘Ensuring that AI research and technology is robust, reliable, trustworthy, and operates within secure constraints.’

⁸² Article 9(4) of the proposed regulation.

⁸³ Articles 19 and 43 of the proposed regulation.

⁸⁴ Article 60(2) of the proposed regulation.

⁸⁵ Articles 11–12 of the proposed regulation.

⁸⁶ Human oversight can be either built into AI or measures can be merely identified so that users can appropriately implement them, according to Article 14(3) of the proposed regulation. Oversight should enable users to understand and monitor AI, interpret its output, decide not to use it, intervene in its operation, and prevent automation bias (Article 14(4)).

⁸⁷ See Eleven Guiding Principles on Lethal Autonomous Weapons (n 19); note that ‘meaningful human control’ is not mentioned as a requirement for autonomous weapons systems in these guiding principles.

⁸⁸ See the discussion of bias above.

⁸⁹ See the discussion of transparency above.

⁹⁰ But see Trusilo and Burri, ‘Ethical AI’ (n 16).

⁹¹ See, for instance, USACM, ‘Algorithmic Transparency’ (n 12) principle no 6: ‘Auditability: Models, algorithms, data, and decisions should be recorded so that they can be audited in cases where harm is suspected.’ (Emphasis removed.)

Overall, the proposed regulation offers compelling evidence of norm diffusion, at least to the extent that the regulation reflects ethical norms on AI. It addresses the three most pressing concerns related to AI of the machine learning type, namely bias due to input data, opacity that hampers predictability and explainability, and vulnerability to misuse (gaming, etc.).⁹² In addressing these concerns, the proposed regulation remains relatively lean. It notably refrains from taking on broader concerns with which modern AI is often conflated, namely dominant market power,⁹³ highly stylized concepts,⁹⁴ and the general effects of technology.⁹⁵

However, the proposed regulation does not fully address the main concerns concerning AI, namely bias and opacity, head-on. It brings to bear a gentle, procedural approach on AI by addressing bias indirectly through data governance and transparency and remedying opacity through interpretability. It entrusts providers with the management of the risks posed by AI and with the judgement of what is tolerable. Providers consequently bear soft duties. In relying on soft duties, the regulation extends the life of ethical norms and continues their approach of indulgence. It thus incorporates the character of ethical norms that lack the commitment of hard law.

On the one hand, it may be unexpected that ethical norms live on to a certain extent, given that the new law on AI is laid down in a directly applicable, binding Union regulation. On the other hand, this is not all that surprising because a horizontal legislative act that regulates all kinds of AI in one go is necessarily less specific on substance than several sectorial acts addressing individual applications. (Though the adoption of several sectorial acts would have had other disadvantages.) Yet, this approach of the proposed regulation begs the question of whether it can serve as a basis for individual, private rights: will natural persons, market competitors, etc. be able to sue providers of high-risk AI for violation of the procedural, managerial obligations incumbent on them under the regulation?⁹⁶

IV. INTERNATIONAL LAW SIDELINED

It is not the case that international law has ignored the rise of AI, while ethics filled the void and laid down the norms. International law – especially the soft type – and ethical principles overlap and are not always easily distinguishable. Yet, even international soft law has been lagging behind considerably. It took until late spring 2019 for the Organization for Economic Co-Operation and Development (OECD) to adopt a resolution spelling out five highly abstract principles on AI.⁹⁷ While the principles address opacity (under transparency and explainability)

⁹² The risk of a responsibility gap is not addressed by the proposed regulation, but by a revision of the relevant legislation on liability, see p 5 of the Explanatory Memorandum to the proposed regulation.

⁹³ See A Ezrachi and ME Stucke, *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy* (2016).

⁹⁴ Bostrom, 'Superintelligence' (n 8); J Dawes, 'Speculative Human Rights: Artificial Intelligence and the Future of the Human' (2020) 42 *Human Rights Quarterly* 573.

⁹⁵ For a broader perspective on AI, see K Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (2021).

⁹⁶ Note the broad geographical scope of the proposed regulation. It applies when providers bring AI into circulation in the Union, but also when output produced outside of the Union is used in it (see Article 2(1)(a) and (c) of the proposed regulation). The substantive scope of the proposed regulation is not universal, though, for it, for instance, largely excludes weapons and cars (see Article 2(2) and (3) of the proposed regulation).

⁹⁷ OECD Recommendation OECD/LEGAL/0449 of 22 May 2019 of the Council on Artificial Intelligence (hereafter OECD, 'Recommendation on AI'; the five principles are the following: 1. Inclusive growth, sustainable development and well-being; 2. Human-centred values and fairness; 3. Transparency and explainability; 4. Robustness, security and safety; 5. Accountability. Another five implementing recommendations advise specifically States to: invest in AI research and development; foster a digital ecosystem; shape the policy environment for AI, including by way of

and robustness (including security and safety), they ignore the risk of bias. Instead, they only generically refer to values and fairness. When the OECD was adopting its non-binding resolution, the European Commission's White Paper⁹⁸ was already in the making. As the White Paper, the OECD Resolution recommended a risk-based approach.⁹⁹ Additionally, the OECD hosts a recent political initiative, the Global Partnership on Artificial Intelligence,¹⁰⁰ which has produced a procedural report.¹⁰¹

Regional organizations have been more alert to AI than universal organizations. Certain sub-entities of the Council of Europe notably examined AI in their specific purview. In late 2018, a commission within the Council of Europe adopted a set of principles governing AI in the judicial system;¹⁰² in the Council of Europe's data protection convention framework, certain principles focussing on data protection and human rights were approved in early 2019.¹⁰³ On the highest level of the Council of Europe, the Committee of Ministers recently adopted a recommendation,¹⁰⁴ which discussed AI ('algorithmic

experimentation; build human capacity and prepare for labour market transformation; and cooperate internationally, namely on principles, knowledge sharing, initiatives, technical standards, and metrics; see also S Voeneky, 'Key Elements of Responsible Artificial Intelligence – Disruptive Technologies, Dynamic Law' (2020) 1 *Ordnung der Wissenschaft* 9, 16.

⁹⁸ White Paper on AI (n 52).

⁹⁹ OECD 'Recommendation on AI' (n 97) point 1.4.c.

¹⁰⁰ OECD, 'OECD to Host Secretariat of New Global Partnership on Artificial Intelligence' (OECD, 15 June 2020) <https://www.oecd.org/newsroom/oecd-to-host-secretariat-of-new-global-partnership-on-artificial-intelligence.htm>; the idea of this initiative may be to counterweigh China in AI: J Delcker, 'Wary of China, the West Closes Ranks to Set Rules for Artificial Intelligence' (*Politico*, 7 June 2021) www.politico.eu/article/artificial-intelligence-wary-of-china-the-west-closes-ranks-to-set-rules/. The OECD initiative is not to be confused with the Partnership on Artificial Intelligence, see Partnership on AI, 'Tenets' (n 28).

¹⁰¹ The Global Partnership on Artificial Intelligence, Responsible Development, Use and Governance of AI, Working Group Report (GPAI Summit Montreal, November 2020) www.gpai.ai/projects/responsible-ai/gpai-responsible-ai-wg-report-november-2020.pdf.

¹⁰² European Commission for the Efficiency of Justice (CEPEJ), 'European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment' (*Council of Europe*, 3-4 December 2018) <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808ff699c>. In sum, it suggested the following guidelines: 1. Ensure compatibility with human rights; 2. Prevent discrimination; 3. Ensure quality and security; 4. Ensure transparency, impartiality, and fairness; 5. Make AI accessible, understandable, and auditable; 6. Ensure user control.

¹⁰³ Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, 'Guidelines on Artificial Intelligence and Data Protection (Council of Europe Convention 108)' (25 January 2019) T-PD(2019)01. The guidelines distinguish between general principles (i), principles addressed to developers (ii), and principles addressed to legislators and policy makers (iii). In summary, the principles are the following: i) 1. Respect human rights and dignity; 2. Respect the principles of Convention 108+: lawfulness, fairness, purpose specification, proportionality of data processing, privacy-by-design and by default, responsibility and demonstration compliance (accountability), transparency, data security and risk management; 3. Avoid and mitigate potential risks; 4. Consider functioning of democracy and social/ethical values; 5. Respect the rights of data subjects; 6. Allow control by data subjects over data processing and related effects on individuals and society. ii) 1. Value-oriented design; 2. Assess, precautionary approach; 3. Human rights by design, avoid bias; 4. Assess data, use synthetic data; 5. Risk of decontextualised data and algorithms; 6. Independent committee of experts; 7. Participatory risk assessment; 8. Right not to be subject solely to automated decision making; 9. Safeguard user freedom of choice to foster trust, provide feasible alternatives to AI; 10. Vigilance during entire life-cycle; 11. Inform, right to obtain information; 12. Right to object. iii) 1. Accountability, risk assessment, certification to enhance trust; 2. In procurement: transparency, impact assessment, vigilance; 3. Sufficient resources for supervisors. 4. Preserve autonomy of human intervention; 5. Consultation of supervisory authorities; 6. Various supervisors (data, consumer protection, competition) should cooperate; 7. Independence of committee of experts in ii.6; 8. Inform and involve individuals; 9. Ensure literacy. See also Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, 'Guidelines on Facial Recognition (Convention 108)' T-PD(2020)03rev4.

¹⁰⁴ Recommendation CM/Rec(2020)1 of 8 April 2020 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems, Council of Europe Committee of Ministers, (hereafter 'Recommendation on

systems',¹⁰⁵ as it calls it) in depth from a human rights perspective. The recommendation drew the distinction between high-risk and low-risk AI that the proposed Union regulation also adopted.¹⁰⁶ It, in large parts, mirrors the European Union's approach developed in the White Paper and the proposed regulation. This is not surprising given the significant overlap in the two organizations' membership.

On the universal level, processes to address AI have moved at a slower pace. The United Nations Educational, Scientific and Cultural Organization is only now discussing a resolution addressing values, principles, and fields of action on a highly abstract level.¹⁰⁷ The United Nations published a High-Level Report in 2019,¹⁰⁸ but it dealt with digital technology and its governance from a general perspective. Hence, the values it lists¹⁰⁹ and the recommendations it makes¹¹⁰ appear exceedingly abstract from an AI point of view. The three models of governance suggested in the report, however, break new ground.¹¹¹

In a nutshell, most of the international law on AI arrives too late. Domestic implementation of ethical norms is already in full swing. Legislative acts, such as the proposed regulation of the EU, are already being adopted. Court and administrative cases are being decided. Meanwhile,

the human rights impacts'). The recommendation is a detailed text that first addresses states and then private actors. After elaborating on scope and context (part A paras 1–15, discussing, for example, synthetic data [para 6], the fusion of the stages of development and implementation of AI [para 7], the presence of both private and public aspect in many algorithmic systems [para 12], and a precautionary approach [para 15]), it lists obligations of states in part B, including data management (para 2), testing (paras 3.3–5), transparency and remedies (para 4), and precautionary measures (para 5, including standards and oversight). These obligations are then tailored to the situation of private actors on the basis of the due diligence approach applicable to business. The obligations in this part are less stringent; see, for instance, the duty to prevent discrimination in para C.1.4.

¹⁰⁵ Recommendation on the human rights impacts (n 104) para A.2.

¹⁰⁶ Recommendation on the human rights impacts (n 104) para A.11.

¹⁰⁷ See UNESCO, 'Draft text of the Recommendation on the Ethics of Artificial Intelligence' SHS/IGM-AIETHICS/2021/APR/4 (UNESCO Digital Library, 31 March 2021) <https://unesdoc.unesco.org/ark:/48223/pf0000376713>; see also UNESCO, 'Artificial Intelligence for Sustainable Development: Challenges and Opportunities for UNESCO's Science and Engineering Programmes' SC/PCB/WP/2019/AI (UNESCO Digital Library, August 2019); see F Molnár-Gábor, Die Herausforderung der medizinischen Entwicklung für das internationale soft law am Beispiel der Totalsequenzierung des menschlichen Genoms, (2012) 72 *Zeitschrift für ausländisches öffentliches Recht und Völkerrecht* 695, for the role of soft law created by UNESCO.

¹⁰⁸ UN High Level Panel on Digital Cooperation, 'The Age of Digital Interdependence: Report of the UN Secretary-General's High-Level Panel on Digital Cooperation' (UN, June 2019) (hereafter 'The Age of Digital Interdependence').

¹⁰⁹ The Age of Digital Interdependence (n 108) 7: Inclusiveness, respect, human-centredness, human flourishing, transparency, collaboration, accessibility, sustainability, and harmony. That 'values' are relative in AI becomes evident from the key governance principles the Report lays down in Section VI. The principles, each of which is explained in one sentence, are the following: Consensus-oriented; Polycentric; Customised; Subsidiarity; Accessible; Inclusive; Agile; Clarity in roles and responsibility; Accountable; Resilient; Open; Innovative; Tech-neutral; Equitable outcomes. Further key functions are added: Leadership, Deliberation; Ensuring inclusivity; Evidence and data; Norms and policy making; Implementation; Coordination; Partnerships; Support and Capacity development; Conflict resolution and crisis management. This long list that appears like the result of a brainstorming begs the question of the difference between the 'values' of the Report on page 7 and the 'principles' ('functions') on page 39 and how they were categorized.

¹¹⁰ The Age of Digital Interdependence (n 108) 29–32; the recommendations include: 1B: Creation of a platform for sharing digital public goods; 1C: Full inclusion for women and marginalized groups; 2: Establishment of help desks; 3A: Finding out how to apply existing human rights instruments in the digital age; 3B: Calling on social media to work with governments; 3C: Autonomous systems: explainable and accountable, no life and death decisions, non-bias; 4: Development of a Global Commitment on Digital Trust and Security; 5A: By 2020, create a Global Commitment for Digital Cooperation; welcoming a UN Technology envoy.

¹¹¹ The Age of Digital Interdependence (n 108) 23–26: The three governance models that are proposed are the following: i) a beefed-up version of the existing Internet governance forum; ii) a distributed, multi-stakeholder network architecture, which to some extent resembles the status quo; and iii) an architecture that is more government driven, while it focuses on the idea of 'digital commons'.

standardization organizations are enacting the technical – and not-so-technical – details. Still, the international law on AI, all of which is soft (and hence not always distinguishable from ‘ethical norms’), is far from being useless. The Council of Europe’s recommendation on algorithmic systems¹¹² added texture and granularity to the existing ethical norms. Instruments that may eventually be adopted on the universal level may spread norms on AI across the global south and shave off some of the Western edges the norms (and AI itself) currently still carry.¹¹³

However, the impact of the ethical norms on AI is more substantial than international legal theory suggests. The ethical norms were consolidated outside of the traditional venues of international law. By now, they are diffusing into domestic law. International law is a bystander in this process. Even if the formation of formally binding international law on AI were attempted at some point,¹¹⁴ a substantial treaty would be hard to achieve as domestic legislatures would have locked in legislation by then. A treaty could only re-enact a consensus established elsewhere, in other words, in ethical norms and domestic law, which would reduce its compliance pull.

V. CONCLUSION AND OUTLOOK

This chapter explained how ethical norms on AI came into being and are now absorbed by domestic law. The European Union’s new proposal for a regulation on AI illustrated this process of ‘bottom-to-bottom’ norm diffusion. While soft international law contributed to forming ethical norms, it neither created them nor formed their basis in a formal, strict legal sense.

This chapter by no means suggests that law always functions or is created in the way illustrated above. Undoubtedly, international law is mainly formed top-down through classical sources. In this case, it also exercises compliance pull. However, in domains such as AI, where private actors – including multinational companies and transnational or domestic non-governmental organizations – freely shape the landscape, a transnational process of law creation takes place. States in such cases tend to realize that ‘their values’ are at stake when it is already too late. Hence, states and their traditional way of making international law are sidelined. However, it is not ill will that drives the process of norm diffusion described in this chapter. States are not deliberately pushed out of the picture. Instead, ethical norms arise from the need of private companies and individuals for normative guidance – and international law is notoriously slow to deliver it. When international law finally delivers, it does not set the benchmark but only re-traces ethical norms. However, it does at least serve to make them more durable, if not inalterable.

The discussion about AI in international law has so far been about the international law that should, in a broad sense, govern AI. Answers were sought to how bias, opacity, robustness, etc., of AI could be addressed and remedied through law. However, a different dimension of international law has been left out of the picture so far. Except for the narrow discussion about autonomous weapons systems within CCW, international lawyers have mainly neglected what

¹¹² Recommendation on the human rights impacts (n 104).

¹¹³ See the useful mapping of AI in emerging economies: ‘Global South Map of Emerging Areas of Artificial Intelligence’ (K4A, 9 June 2021) www.k4all.org/project/aiecosystem/; Knowledge for All, a foundation, conducts useful projects on development and AI, see www.k4all.org/project/?type=international-development.

¹¹⁴ The Council of Europe is currently deliberating on whether to draft a treaty on AI: Feasibility Study, Council of Europe Ad Hoc Committee on Artificial Intelligence (CAHAI), CAHAI(2020)23.

AI means for international law itself and the concepts at its core.¹¹⁵ Therefore, the next step to be taken has to include a re-assessment of central notions of international law in the light of AI. The notions of territoriality/jurisdiction, due diligence duties concerning private actors, control that is central to responsibility of all types, and precaution should consequently be re-assessed and recalibrated accordingly.

¹¹⁵ A further dimension relates to the use of AI for international lawyers, see A Deeks, 'High-Tech International Law' (2020) 88(3) *George Washington Law Review* 574–653; M Scherer, 'Artificial Intelligence and Legal Decision-Making: The Wide Open? — A Study Examining International Arbitration' (2019) 36 *Journal of International Arbitration* (5) 539–574; for data analysis and international law, see W Alschner, 'The Computational Analysis of International Law' in R Deplano and N Tsagourias (eds), *Research Methods in International Law: A Handbook* (2021) 204–228.