Multilateral organisations

Anna Yamaoka-Enkerlin

White & Case*

In January 2020, Google Chief Executive Officer (CEO) Sundar Pichai made waves when he declared that 'there is no question in my mind that artificial intelligence needs to be regulated', and called 'international alignment critical'.[1]

Three years later, as discussed at length below, we can take stock and say that progress is underway. But at the same time as the critical need for ethical AI standards is clearer than ever, the prospect of seamless 'global' alignment on AI regulation seems more unlikely than ever.[2]

Events over last three years – from Russia's invasion of Ukraine to the banning of China's Huawei from the 5G networks of many Western countries – have also heightened the sense in which the future may be shaped by a struggle that is as strategic as it is ideological. AI will shape, facilitate, and accelerate this struggle. Although standard setting may convey a sense of neutrality, this disguises an intense ethical, commercial and geopolitical struggle to control the future of AI.[3] Worldwide acceptance of one's proposed standard, especially when that standard tracks a company's proprietary technology, allows that company or country to reap commercial rewards and set the norms for the future development of AI; the emergence of global standards 'not only impacts the power of nation-states, but also changes the power of corporations'.[4]

The aim of this chapter is to highlight briefly some of the most critical intergovernmental AI policy initiatives currently underway. Most deal in high-level, generally applicable principles rather than being tailored to the context of AI use in legal or other professional contexts. But a sense of the multilateral efforts taking place in this area should be relevant to all professionals who have an interest in anticipating the future of technological progress, incoming regulation and possible liability while leveraging the ethical use of AI as a competitive advantage.

[*] Many thanks to Sofya Cherkasova for her research assistance in updating the second edition of this chapter.

[1] Sundar Pichai, 'Why Google thinks we need to regulate AI', *Financial Times* (London, 20 January 2020), see www.ft.com/content/3467659a-386d-11ea-ac3c-f68c10993b04 accessed 2 July 2020.

[2] See EU special committee on Artificial Intelligence in a Digital Age, Artificial Intelligence Diplomacy, June 2021, https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662926/IPOL_STU(2021)662926_EN.pdf; Joseph Bouchard, 'AI Strategic Competition, Norms, and the Ethics of Global Empire', The Diplomat, (Arlington , 1 December 2021) https://thediplomat.com/2021/12/ai-strategic-competition-norms-and-the-ethics-of-global-empire accessed 12 February 2023.

[3] Alan Beattie, 'Technology: How the US, EU and China compete to set industry standards', *Financial Times* (London, 24 July 2019) www.ft.com/content/0c91b884-92bb-11e9-aea1-2b1d33ac3271 accessed 26 July 2020.

[4] Aynne Kokas, 'Cloud Control: China's 2017 Cybersecurity Law and its Role in US Data Standardization', 29 July 2019, see https://ssrn.com/abstract=3427372 or http://dx.doi.org/10.2139/ssrn.3427372 accessed 26 July 2020.

Organisation for Economic Co-operation and Development (OECD)

The OECD's Principles on Artificial Intelligence – the first intergovernmental standards on AI – were adopted by 42 countries on 22 May 2019.[5]

Although these principles are meant to apply across all sectors, the possibility of overlap with other professional regulation is acknowledged by the preamble, which 'underlines' that 'certain existing regulatory and policy frameworks already have relevance to AI, including those related to [...] responsible business conduct'.[6]

Contained within the OECD Council Recommendation on AI, the principles are delivered in two sections. The first section, 'principles for responsible stewardship of trustworthy AI', elaborates on five 'complementary value-based principles':

1. inclusive growth, sustainable development and wellbeing;

2. human-centred values and fairness;

3. transparency and explainability;

4. robustness, security and safety; and

5. accountability.

The second section, 'national policies and international cooperation for trustworthy AI', explicates five 'recommendations' for signatories:

1. investing in AI R&D;

2. fostering a digital ecosystem for AI;

3. shaping an enabling policy environment for AI;

4. building human capacity and preparing for labour market transformation; and

5. international cooperation for trustworthy AI.

The OECD Committee on Digital Economy Policy is responsible for monitoring the implementation of these recommendations, as well as the development of more practical guidance through fostering international dialogue at the OECD AI Policy Observatory.[7]

Although OECD recommendations are not binding, they 'are highly influential', and in the past, have formed the starting point for government negotiations on

[5] OECD, 'OECD Principles on Artificial Intelligence', see www.oecd.org/going-digital/ai/principles accessed 2 July 2020.

[6] OECD, 'Recommendation of the Council on Artificial Intelligence' (2019), see https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 accessed 2 July 2020.

[7] OECD, 'Artificial Intelligence', see www.oecd.org/going-digital/ai accessed 10 July 2020.

national legislation, as seen by the influence of the OECD Privacy Guidelines on privacy legislation worldwide.[8]

The influence of the OECD's recommendations is also instantiated by two other intergovernmental pacts on the responsible development and use of AI: The G20's 'Osaka Leaders' Declaration' and associated initiatives, and the Global Partnership on Artificial Intelligence (GPAI).

The G20

In June 2019, the Group of Twenty (G20) issued the 'Osaka Leaders' Declaration' on the digital economy. Along with pushing for concepts such as cross-border 'Data Free Flow with Trust', the G20 committed to a 'human-centred approach to AI' and welcomed the 'non-binding' G20 AI principles, which are drawn from the OECD principles.[9] In 2021 G20 Digital Minsters issued a Declaration, reaffirming their commitment to these AI principles and issued the 'G20 Policy Examples on How to Enhance the Adoption of AI by MSMEs and Start-ups'.[10]

The Global Partnership on Artificial Intelligence

The Global Partnership on Artificial Intelligence (GPAI) stems from a pledge by Canada and France to bridge the theory and practice of 'a vision of a human-centric artificial intelligence'.[11] GPAI was inspired in part by the Intergovernmental Panel on Climate Change (IPCC) to develop global governance of AI.[12] Founding GPAI parties, including Australia, France, Germany, India, Italy, Mexico, Singapore, Slovenia, South Korea, the United Kingdom, the United States, and the European Union, have pledged to 'support the responsible and human-centric development and use of AI in a manner consistent with human rights, fundamental freedoms, and our shared democratic values, as elaborated in the OECD Recommendation on AI'.[13]

[8] OECD, 'OECD Principles on Artificial Intelligence', see www.oecd.org/going-digital/ai/principles accessed 10 July 2020.

[9] Government of Canada, Global Affairs, 'G20 Osaka Leaders' Declaration', see https://www.international.gc.ca/world-monde/international_relations-relations_internationales/g20/2019-06-29-g20_leaders-dirigeants_g20.aspx? accessed 29 June 2019.

[10] G20, 'Declaration of G20 Digital Ministers: Leveraging Digitalisation for a Resilient, Strong, Sustainable and Inclusive Recovery', 5 August 2021, see http://www.g20.utoronto.ca/2021/210805-digital.html accessed 28 June 2022.

[11] 'Innovation, Science and Economic Development Canada', Joint Statement from Founding Members of the Global Partnership on Artificial Intelligence, see https://www.canada.ca/en/innovation-science-economic-development/news/2020/06/joint-statement-from-founding-members-of-the-global-partnership-on-artificial-intelligence.html accessed 14 June 2020.

[12] Nicolas Miailhe, 'Why We Need an Intergovernmental Panel for Artificial Intelligence', *Our World*, 21 December 2018, see https://ourworld.unu.edu/en/why-we-need-an-intergovernmental-panel-for-artificial-intelligence accessed 14 June 2020.

[13] See n 8 above.

Hosted by the OECD in Paris, GPAI has focused its initial efforts on four working group themes:

1. Responsible AI – studying the effects of social media recommender systems on users[14] and elaborating on recommendation for government action in the area of climate change and AI.[15]

2. Data governance – producing guidance for policymakers in the sphere of data justice and highlighting the potential of data trusts to address social issues and climate change.[16]

3. The future of work – analysing 'how AI can be used in the workplace to empower workers'.[17]

4. Innovation and commercialisation – examining the adoption of AI by small and medium-sized enterprises (SMEs) and ways to protect AI innovation and intellectual property.[18]

The United Nations

The UN is engaged in AI-related activities across the entire organisation,[19] but the following are stand-out efforts at global coordination to secure the beneficial use of AI, in particular to achieve the Sustainable Development Goals (SDGs).

International Telecommunications Union (ITU)

The ITU is a specialised UN agency for information and communications technology (ICT). A public-private membership that includes 193 Member States and over 900 companies, universities, and international and regional organisations, its functions include developing ICT policies and internationally interoperable technical standards.

Although two private regulatory standard networks – the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) – are the leading bodies for standard setting in digital technologies, the ITU is the only

[14] GPAI, 'Responsible AI for Social Media Governance: A proposed collaborative method for studying the effects of social media recommender systems on users' (November 2021), see https://gpai.ai/projects/responsible-ai/social-media-governance/responsible-ai-for-social-media-governance.pdf accessed 28 June 2022.

[15] GPAI, 'Climate change and AI: Recommendations for government action' (November 2021), see https://gpai.ai/projects/responsible-ai/environment/climate-change-and-ai.pdf accessed 28 June 2022.

[16] GPAI, Working Group on Data Governance, see https://gpai.ai/projects/data-governance accessed 28 June 2022.

[17] GPAI, Working Group on the Future of Work, see https://gpai.ai/projects/future-of-work accessed on 28 June 2022.

[18] GPAI, Working Group on Innovation and Commercialization, see https://gpai.ai/projects/innovation-and-commercialization accessed 28 June 2022.

[19] ITU, 'United Nations Activities on Artificial Intelligence (AI)' (2021), see https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-UNACT-2021-PDF-E.pdf accessed 28 June 2022.

treaty-based organisation with Member States.[20] To a greater degree than ISO, IEC, and prominent industrial associations and consortia such as the Institute of Electrical and Electronics Engineers (IEEE)[21], the ITU's standards are notable for being driven by corporate and national interests outside of North America and the EU. The standards that it produces are particularly influential in the developing world.[22]

Relevant ITU focus groups include the ITU Group on Machine Learning for Future Networks and on AI for Autonomous and Assisted Driving.[23] In line with China's strategy to become the world's standards supplier,[24] Chinese companies have been particularly active in the ITU, gaining acceptance for their standards proposals in the areas of facial recognition and other types of visual surveillance.[25] The ITU also convenes the AI for Good Global Summit, the 'leading UN platform for global and inclusive dialogue on AI', which collaborates with public and private bodies, as well as over 37 UN agencies to 'identify strategies to ensure that AI technologies are developed in a trusted, safe and inclusive manner, with equitable access to their benefits'.[26] Finally, it hosts an 'AI repository' to gather information on AI-related projects that aim to advance progress on the UN SDGs.

UN Educational, Scientific and Cultural Organization (UNESCO)

On 24 November 2021 UNESCO adopted the Recommendation on the Ethics of Artificial Intelligence, 'the first global standard-setting instrument on the ethics of artificial intelligence'.[27] A first draft of the Recommendation was proposed by an Ad-Hoc Expert Group for the Recommendation on the Ethics of AI composed of 24 specialists in AI and Ethics,[28] and was then developed after a consultation process that included: (1) public online consolation; (2) Regional and Sub-regional

[20] Jeffrey Deng, 'Balancing Standards: U.S. and Chinese Strategies for Developing Technical Standards in AI', *NBR*, 1 July 2020, www.nbr.org/publication/balancing-standards-u-s-and-chinese-strategies-for-developing-technical-standards-in-ai accessed 10 July 2020.

[21] For an important contribution to the development of ethical AI standards with recommendations for implementation developed by over 700 global experts, see Kyarash Shahriari and Mana Shahriari, 'Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems', IEEE, 2017, https://ieeexplore.ieee.org/document/8058187 accessed 12 February 2023.

[22] Anna Gross, Madhumita Murgia and Yuan Yang, 'Chinese tech groups shaping UN facial recognition standards' *Financial Times* (London, 1 December 2019), see www.ft.com/content/c3555a3c-0d3e-11ea-b2d6-9bf4d1957a67 accessed 10 July 2020.

[23] ITU, 'International Standards for an AI-Enabled Future', *ITU News*, 6 July 2020, see https://news.itu.int/international-standards-for-an-ai-enabled-future accessed 10 July 2020.

[24] Matt Sheehan, Marjory Blumenthal And Michael R Nelson, 'Three Takeaways From China's New Standards Strategy', *The Carnegie Foundation*, 28 October 2021, https://carnegieendowment.org/2021/10/28/three-takeaways-from-china-s-new-standards-strategy-pub-85678 accessed 12 February 2023.

[25] See n 22 above; see also Asia Society Policy Institute, 'Stacking the Deck: China's Influence in Digital Rules Setting', 30 November 2021, https://asiasociety.org/policy-institute/events/stacking-deck-chinas-influence-digital-rules-setting accessed 12 February 2023.

[26] AI for Good Global Summit 2020, see https://aiforgood.itu.int accessed 10 July 2020.

[27] UNESCO, 'Recommendation on the Ethics of Artificial Intelligence', see https://unesdoc.unesco.org/ark:/48223/pf0000381137 accessed 28 June 2022.

[28] UNESCO, 'Composition of the Ad Hoc Expert Group (AHEG) for the Recommendation on the Ethics of Artificial Intelligence' https://unesdoc.unesco.org/ark:/48223/pf0000372991 accessed 12 February 2023.

consultations co-organised with host countries around the world, and (3) multi-stakeholder workshops in 25 countries.[29] The Recommendation, which was endorsed by 193 countries, 'aims to provide a basis to make AI systems work for the good of humanity'.[30] It establishes the values that serve as benchmark for any AI system: respect, protection and promotion of human rights, environment and ecosystem flourishing, ensuring diversity and inclusiveness, living in peaceful, just and interconnected societies. Building on these values, the Recommendation outlines 11 areas of policy action: Ethical Impact Assessment, Ethical Governance and Stewardship, Data Policy, Development and International Cooperation, Environment and Ecosystems, Gender, Culture, Education and Research, Communication and Information, Economy and Labour, Health and Social Well-Being.

UN Convention on Certain Conventional Weapons (CCW)

States which are parties to the UN Convention on Certain Conventional Weapons (CCW) have been discussing the regulation of emerging lethal autonomous weapons systems (LAWS), with the UN Secretary-General repeatedly calling on states to conclude a new relevant international treaty.[31] In 2017, a Group of Governmental Experts was established to assess emerging legal questions related to LAWs. In 2019, at the recommendation of the Group, the 2019 Meeting of the High Contract Parties to the CCW adopted 11 guiding principles on LAWs.[32] These principles: affirm the applicability of international law – including international humanitarian law – to the development, acquisition, and deployment of LAWs; highlight the need to consider the risks of proliferation, including acquisition by terrorist groups; and call for retaining human responsibility and accountability across the entire life cycle of the weapons systems – all while recognising the need to balance military necessity and humanitarian considerations. But apart from the publication of these principles, substantive progress on a binding international treaty has been stalled by opposition from military powers such as China, Russia, the UK and the US.[33]

[29] UNESCO, Recommendation on the Ethics, 'Ethics of Artificial Intelligence', https://en.unesco.org/artificial-intelligence/ethics#recommendation accessed 13 September 2022.

[30] *Ibid*.

[31] 'Autonomous weapons that kill must be banned, insists UN Chief', *UN News*, 25 March 2019, see https://news.un.org/en/story/2019/03/1035381 accessed 10 July 2020.

[32] UN, 'background on LAWS at the CCW' https://www.un.org/disarmament/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw accessed 18 September 2022.

[33] Zelin Liu, and Michael Moodie, 'International Discussions Concerning Lethal Autonomous Weapon Systems', see Reuters, 'U.N. talks adjourn without deal to regulate 'killer robots' https://www.reuters.com/world/un-talks-adjourn-without-deal-regulate-killer-robots-2021-12-17; US Congressional Research Service, 'International Discussions Concerning Lethal Autonomous Weapon Systems', 21 December 2021 https://sgp.fas.org/crs/weapons/IF11294.pdf accessed 12 February 2023.

UN Centre for Artificial Intelligence and Robotics (UNICRI)

Launched in 2015, UNICRI's aim is to 'enhance understanding of the risk-benefit duality of Artificial Intelligence and Robotics through improved coordination, knowledge collection and dissemination, awareness-raising and outreach activities'.[34] UNICRI has partnered with INTERPOL to study the impact of AI in law enforcement and to develop the 'Toolkit for Responsible Artificial Intelligence Innovation in Law Enforcement', which is expected to be presented to experts in late 2022.[35] UNICRI has also launched the AI for Safer Children initiative, and has worked with the UN Counter-Terrorism Centre to analyse the use of AI in counter-terrorism activities.

The European Union

The European Union has been prolific in its development of AI policy initiatives, in part because the absence of a common EU framework for addressing the challenges posed by AI risks fragmenting its internal market. The EU's AI policy development is included in this chapter on multi-lateral initiatives because EU technology policy precedent affects not only all the EU Member States but also has proved highly influential globally.

In April 2021 the EU Commission launched its proposal for a 'Regulation for Laying Down Harmonised Rules on Artificial Intelligence' (the AI Act) following various EU policy initiatives focused on Ethical AI. One of the stated aims of the AI Act is to 'position […] Europe to play a leading role globally'. The EU has recognised as a priority 'the need to act as a global standard-setter in AI', explicitly recognising that falling behind in the race for global tech leadership will leave room for the adoption of standards developed in non-democratic countries to dominate.

On 6 December 2022, the European Council – a body that includes ministers from each EU Member State – finalised its modifications to the EU Commission's proposal (compromise text). As of the time of writing, the next step is for the European Parliament to adopt its own position. To that end, the Parliament is reportedly working through 3,300 proposed amendments, many focused on the definition of AI, the high-risk categorisation of certain AI systems (discussed below) and the governance scheme that the Commission's AI Act proposes. Once the Parliament finalises its own position, the Council of the EU and the European Parliament are likely to hold negotiations with the EU Commission (the 'trilogue') before a final text is decided on and adopted by both the Council and the Parliament. The AI Act will is likely to be passed in early 2024. The main features of the Act are summarised below.

[34] UNICRI Centre for Artificial Intelligence and Robotics, The Hague, https://unicri.it/in_focus/on/unicri_centre_artificial_robotics, accessed 10 July 2020.

[35] UNICRI, 'UNICRI and INTERPOL formally kick-off next phase of work on Toolkit for Responsible AI Innovation in Law Enforcement with funding from the European Commission', 29 November 2021, see https://unicri.it/News/Toolkit-AI-Law-Enforcement-INTERPOL-EC-UNICRI, accessed 28 June 2022.

The definition of AI

One of the more contentious aspects of the Act is the definition of AI to be adopted. The challenge is deciding on a definition not only captures the range of the 'high-risk' AI that the EU wants to target and is flexible enough to accommodate new AI techniques, but also avoids being so over inclusive as to impose undue burdens on innovation. The Commissions' draft Act defines AI systems as a 'software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with'. Annex I features three categories: machine learning approaches, logic and knowledge-based approaches, and statistical approaches. Finally, AI systems that are developed exclusively for use by the military (itself a contentions term) are excluded from the Act – a move which may seem out of step with the thrust of the AI Act as an instrument expounding AI ethics, leaving a gap that remains to be filled.

The European Council, concerned that the AI's definition unnecessarily captures many software applications that should not be burdened with compliance under the Act, has proposed a definition which is narrower in some respects, but which also removes the explicit reference to humans in 'human-defined set of objectives.' The Council has also commissioned a group to study the definition of AI relation to general purpose AI (GPAI). Opinion remains divided as to how the AI act should approach the unique opportunity, and regularity challenges, that GPAI poses. Finally, the Council has proposed adding a clause that excludes, along with AI developed exclusively for military purposes, AI that is developed exclusively for national security purpose.

Subjects of regulation

The proposed AI Act imposes new duties on various players in the 'AI value chain' including AI 'providers' who place on the market or put into service AI systems within the EU – irrespective of their locations – provided that 'the output produced by the system is used in the EU.' Providers can be natural or legal persons that are public or private. AI users will also incur duties under the act, except when using an AI system in the course of a non-professional activity. Under both the Commissions' Draft AI act and the Council's comprise text, importers and distributors will be treated as 'providers' under the legislation if, among other circumstances, they place on the market a high-risk AI system under their name, if they modify the purpose of a high-risk AI system that is already placed on the market or put into service, or they make a substantial modification to the high-risk system. In these cases, the original provider will be relieved of its obligations as a provider.

The proposed extraterritorial dimension of the AI Act – applying as it does to any provider or user so long as the relevant AI system 'output produced by those system is used in the EU' may, like similar provisions in the General Data Protection Regulation (GDPR), help to drive a 'Brussels Effect'. The Brussel's Effect refers to the way regulatory globalisation is caused by the extraterritorial influence of EU law. The GDPR achieved the 'Brussels Effect' through the territoriality provisions of its Article 3, which clarify that the GDPR's provisions apply to the processing of personal data of data subjects who are in the EU by a controller or processor not established in the EU. Furthermore, by conditioning personal data law transfers out of the EU on an 'adequacy' assessment – where 'adequate' means 'essentially equivalent' – the EU has secured leverage to demand that its international trading partners replicate its policy vision. Many jurisdictions have taken the GDPR as a starting point for designing their own legislation.

EU lawmakers have referred to the Brussels Effect as a reason to pass the AI Act quickly, although opinion is split as to whether, or how far, the EU will achieve this effect. Arguably we are already seeing tangible examples of the Brussel's Effect on AI regulation in action, as reflected in its influence on Brazil's forthcoming AI legislation.

Duties imposed

To achieve the EU's aim of fostering innovation and protecting EU values and fundamental rights, the European Commission adopted a 'risk' based approach, meaning that different levels of regulation will be applied depending on the level of risk that an AI system is considered to pose to individuals and society.

First, there are AI applications and systems that are considered under the Act to create unacceptable risks of violating EU values or fundamental rights. This includes subliminal manipulation resulting in physical or psychological harm, exploiting children or mentally disabled persons resulting in physical or psychological harm, general purpose social scoring, and remote biotitic identification by law enforcement in publicly accessible spaces (with exceptions.)

Second, there are AI applications and systems that are considered 'high-risk'. Under the Commission's proposal, whether an AI application or system is 'high risk' is determined based on the 'intended purpose of the system and on the severity of the possible harm and the probability of its occurrence'. Examples of 'high-risk' applications fall into two groups: AI involved in safety components of regulated products (eg, medical devices) which are subject to third-party assessment under the relevant sectoral legislation; and certain standalone systems that fall under various categories. The proposed categories include law enforcement, management and operation of critical infrastructure, education and vocational training, employment and worker management, migration and asylum, access to essential private services and public services, and administration of justice.

Given the extensive regulation of ‘high-risk’ AI, and the associated costs there is considerable debate about the scope of this category. The European Commission’s impact assessment proposes that only five-to-15 per cent of currently available AI applications are ‘high-risk’ under the draft regulation. That number may increase if the EU more directly targets GPAI, where much research, product development – and hype for the future of AI – currently lies.

AI that is considered ‘high risk’ will only be allowed to be marketed by a provider if the providers conforms with a suite of legal requirements, including, but not limited to: the use of high-quality datasets; data and record-keeping to enhance traceability; the adoption of human oversight measures and implementation of high standards of algorithmic interpretability; accuracy; robustness; and cybersecurity, as well as technical documentation demonstrating compliance.

To govern the regulation of ‘high-risk’ AI, the Act also introduces a mandatory certification system. Under the draft regulation, before placing a high-risk system on the market, AI providers must ensure that the design and development of the system complies with the AI regulation, perform a conformity assessment to document this compliance, notify national authorities that will be tasked with administering an AI certification scheme, and then obtain a certification.

The AI Act also registers importers as enforcers, requiring that importers ensure that the relevant creator/provider of the high-risk AI system that the importer intends to place on the market has carried out a conformity assessment and obtained the required certification. If the importer finds that the creator/provider of the high-risk AI system is non-compliant, they must refuse to place the system into the market and, where there is a risk that this AI will be introduced to the market even if the importer refuses to do so, the importer must inform the AI provider and the market authorities.

The third category addressed by the proposed regulation is AI activity perceived to present a lower level risk, which will be subject only to minimal transparency requirements. Transparency obligations will apply for systems that: (1) interact with humans; (2) are used to detect emotions or determine association with (social) categories based on biometric data; or (3) ‘deep fakes’, which are defined as audio or video content that ’appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful generate or manipulate content (“deep fakes”).’ The requirement to label deep fakes does not apply to deep fakes authorised by law enforcement or where relevant rights, such as the freedom of expression guaranteed in the Charter of Fundamental Rights of the EU – leaving the scope of this requirement still unclear.

The fourth category of AI applications perceived to present low or no risk has no mandatory requirements although voluntary compliance will be encouraged.

Penalties

Individuals and companies who violate the act by, among other things, engaging in forbidden practices, failing to meet their obligations for high-risk systems, or not cooperating with the competent national authorities will be subject to penalties. Under the legislation, the penalty incurred will depend on the type of violation, and the identity of the party that commits the violation (ie, whether they are a provider, importer, distributer, user, etc and, where relevant, the size of the company found to be infringing the Act). The most severe fines would be levied for breaches of the ban on AI systems that pose an unacceptable risk (such as creating a social scoring system) which can reach a maximum of €30m or six per cent of the violator's annual revenue. Companies which fail to meet their obligations with regards to High-Risk AI will face fines of up to €20m or four per cent of their total annual revenue. For small and medium-sized enterprises (SMEs) and start-ups, the fines can be up to two per cent of their annual revenue.

Fostering innovation

As part of the EU's commitment to fostering innovation, and avoiding undue burdens imposed by the Act, the proposed regulation includes provisions for the creation of regulatory sandboxes, which are testing grounds for AI applications that operate under specific, limited conditions. These sandboxes, which start-ups and SMEs would be given privileges access to provided they meet certain criteria, would be used to foster innovation by allowing companies and researchers to test and develop new AI technologies in a controlled environment, without the full burden of compliance with all existing regulations. The idea, which is already being piloted , is to provide a safe space for experimentation, learning, and development of best practices, while still protecting the public interest and ensuring that AI is used in a responsible manner.

The regulatory framework has attracted various criticisms, including for undue vagueness and insufficient regulation of algorithmic fairness. Another challenge going forward – which may ultimately be the key to the Act's success or failure – will be operationalising the Act's Requirements and distilling them into technical standards; a task already being taken up by EU standard setting organisations. Even so, it is a ground-breaking regulation that is already affecting AI deployment and accelerating discussions about ethical AI worldwide.

Conclusion

The intergovernmental efforts described above could genuinely be criticised as overly vague 'ethics-washing',[36] with minimal substantive influence on design – not least because by the time regulations and standards have been finalised, they may well be out of date. At worst, one might think that attempts to regulate AI will inevitably have a stifling effect on technological progress. Others think that AI policy is best left to the private sector alone. But Pichai, at least, would appear to disagree. Without minimising the considerable work that needs to be done in operationalising these myriad principles and developing ways to verify compliance, even these high-profile efforts should not be simply dismissed. It is not only the end-result, but also the process – in particular sharing and testing of ideas across silos that accompanies these regulatory efforts – which itself advances progress towards ethical AI.[37] We have also seen in the past how 'soft law' has led to transformed 'hard law', as with the influence of the OECD privacy principles on privacy legislation around the world, as well as how ethical considerations are affecting the development of technical standards. In an area as economically and geopolitically fraught as the future of AI development, cooperation towards the mission of steering AI embodied in these multilateral efforts is cause for optimism.

[36] Karen Hao, 'In 2020, let's stop AI ethics-washing and actually do something', *MIT Technology Review*, 27 December 2019, see www.technologyreview.com/2019/12/27/57/ai-ethics-washing-time-to-act accessed 2 July 2020.

[37] eg, although the AI Act has not yet passed, researchers at the University of Oxford are already using available information to develop a conformity assessment procedure for AI systems, see, Luciano Floridi et al, 'capAI – A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act', SSRN, 23 March 2022, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4064091 accessed 12 February 2023.