DAFTAR PUSTAKA

Buku:

Abdul Kadir & Terra Ch. Triwahyuni. 2013. *Pengantar Teknologi Informasi Edisi Revisi.* Peberbit ANDI. Yogyakarta.

Alina Kaczorowska. 2002. *Textbook: Public International Law*. Old Balley Press. London.

Bambang Permadi dkk, Badan Pengkajian dan Penerapan Teknologi (BPPT), S*trategi Nasional Kecerdasan Artifisial Indonesia 2020 – 2045*, Jakarta, BPPT 2020,

Christoph Bartneck et.al. 2021. *An introduction to ethics in robotics and AI*. Springer Nature Switzerland. Switzerland.

Corinne Cath. 2018. *Governing Artificial Intelligence: Ethical. Legal. and Technical Opportunities and Challenges.* Royal Society. London.

Damian M. Bielicki. 2022. *Regulating Artificial Intelligence in Industry*. Routledge. New York.

Donald R. Franceschetti. 2018. *Principles of Robotics & Artificial Intelligence*. Salem Press & Grey House Publishing. Amenia.

Isaac Asimov. 1950*. "Runaround" I. Robot (The Isaac Asimov Collection)* Doubleday. New York City.

Jacob Turner. 2019. *Robot Rules: Regulating Artificial Intelligence*. Fountain Court Chambers. London.

Jaemin Lee. 2022. *Artificial Intelligence and International Law*. Springer Nature Singapore. Singapore.

Jonaedi Effendi, Johnny Ibrahim. 2016. *Metode Penelitian Hukum Normatif dan Empiris*. Prenada Media Group. Depok.

Marcelo Corrales et.al. 2018. *Robotics. AI. and the Future of Law*. Springer Nature Singapore. Singapore.

Michael White. 2005. *Isaac Asimov: A Life of the Grand Master of Science Fiction*. Carroll & Graf. New York City.

Mochtar Kusumaatmadja. 1982. *Pengantar Hukum Internasional.* Binacipta. Jakarta.

Neil Wilkins. 2019. *AI: An Essential Beginner's Guide to AI, Machine learning. Robotics, The Internet of Things, Neural Networks, Deep Learning*. Bravex Publications.

Nick Bostrom et.al. 2008. *Global Catastrophic Risks*. Oxford University Press. Oxford.

Pat Nakamoto. 2017*. Neural Networks and Deep Learning*. Self-Published. Silicon Valley.

Peter Mahmud Marzuki. 2017. *Penelitian Hukum*. Kencana. Jakarta.

Shin-Yi Peng, Ching-Fu Lin, Thomas Streinz. 2021. *Artificial Intelligence and International Economic Law Distruption*. Cambridge University Press. Cambridge.

Stuart J. Russel, Peter Norvig. 2010. *Artificial Intelligence : A Modern Approach.* Pearson Education. New Jersey.

Suyanto. 2014. *Artificial Intelligence Revisi Kedua*. Informatika. Bandung.

Themistoklis Tzimas. 2021. *Legal and Ethical Challenges Artificial Intelligence from an International Law Perspective*. Springer Nature Switzerland AG. Switzerland.

Widodo Budiharto. 2018. *AI For Beginner*. Bina Nusantara University. Jakarta.

Woodrow Barfield & Ugo Pagallo. 2018. *Research Handbook on the Law of Artificial Intelligence.* Edward Elgar Publishing Limited. Cheltenham.

Yuli Vasiliev. 2020. *Natural Language Processing with Python and SpaCy - A Practical Introduction*. No Starch Press. San Fransisco.

Konvensi, Rancangan Konvensi, & Peraturan Terkait:

Europe Union. *European Charter of Fundamental Rights*. (EUR-Lex-12012P/TXT).

Europe Union - 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 2021/0106 (COD).

OECD. *Recommendation of the Council on Artificial Intelligence*. (OECD/LEGAL/0449).

OHCHR. *International Convenant on Civil and Political Rights (ICCPR).*

OHCHR. Universal Declaration of Human Rights (UDHR)

ORAD Committee. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles J3016*. SAE International. Juni 2018. Amerika Serikat.

NHTSA (National Highway Traffic Safety Administration). *Federal Automated Vehicles Policy – Accelerating the Next Revolution in Roadway Safety*. September 2016. Amerika Serikat.

UNDP. *Future Forward—UNDP Digital Strategy (2019)*

UNDP. *United Nations Development Programme – Digital Strategy 2022-2025*

UNESCO. *UNESCO Draft Recommendation on The Ethics of Artificial Intelligence*. (SHS/IGM-AIETHICS/2021/JUN/3 Rev.2).

UNESCO AI and Education-Guidance for Policy Makers (2021) – United Nations Educational, Scientific and Cultural Organization (UNESCO)

UNICRI (United Nations Interregional Crime and Justice Research Institute). *ArtificiaI Intelligence and Robotics for Law Enforcement*. Juli 2018. Singapore.

UNODA. UN Convention on Certain Conventional Weapons (CCW)

Request for Comments on a Draft Memorandum to the Heads of Executive Departments and Agencies 'Guidance for Regulation of Artificial Intelligence Applications 2020'.

WIPO. Artificial Intelligence and Intellectual Property Policy. Provisional Agenda. WIPO/IP.AI/2/GE/20/INF/1/PROV.2 (Jun. 19. 2020)

WIPO. Revised Issues Paper on Intellectual Property Policy and Artificial Intelligence. IP/AI/2GE/20/1 Rev. (May 21. 2020)

Jurnal Ilmiah:

Abu Ahmad Hania. 'Mengenal Artificial Intelligence. Machine Learning. Neural Network. dan Deep Learning'. Jurnal Teknologi Indonesia. Yayasan Teknologi Indonesia. Jakarta. Juni 2017

Aditya Bhargava. "Research Paper on Artificial Intelligence". International Journal of Social Science and Economic Research. G. D. Goenka Signature School. Issue: 10. Vol. 06. October 2021

Aleks Attanasio et.al. 'Autonomy in Surgical Robotics'. Annual Review of Control. Robotics. and Autonomous Systems. School of Electronic and Electrical Engineering. University of Leeds. Leeds. 2021

Annemarie Bridy. "Coding Creativity: Copyright and the Artificially Intelligent Author." Stanford Technology Law Review. Stanford University. Vol.5 No. 1. Stanford. 2012.

Chris Smith et. al. 'The History of Artificial Intelligence'. University of Washington. Seattle. December 2006

Denise Garcia. Lethal Artificial Intelligence and Change: The Future of International Peace and Security. International Studies Review. Vol. 20. Issue 334. 2018.

Eileen Donahoe. and Megan MacDuffee Metzger. Artificial Intelligence and Human Rights. Journal of Democracy. Vol. 30. no. 2. 2019

Hanif Khan. “Different Types of Artificial Intelligence Systems”. University of Science and Technology of Ha Noi. Hanoi. September 2021.

John. O. McGinnis & Russell G. Pearce. The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services. Fordham Law Review. Fordham University School of Law. Vol. 82. 2014.

Jorma Rantanen. Franklin Muchiri. and Suvi Lehtinen. ‘Decent Work. ILO’s Response to the Globalization of Working Life: Basic Concepts and Global Implementation with Special Reference to Occupational Health’. International Journal of Environmental Research and Public Health. May 12. 2020.

Kajian Kominfo: Big Data. Kecerdasan Buatan, Blockchain, dan Teknologi Finansial di Indonesia. Direktorat Jenderal Aplikasi Informatika Kementerian Komunikasi dan Informatika Republik Indonesia. Jakarta. Juli 2018.

Kamal Hussain. ‘Artificial Intelligence and its Applications Goal’. International Research Journal of Engineering and Technology (IRJET). Department of Information Technology & Mathematic ICFAI University. India. January 2018.

Karen Yeung. ‘Introductory Note to Recommendation of the Council on Artificial Intelligence (OECD)’. The American Society of International Law. Juli 2019.

Karman. ‘Strategi Dalam Mengembangkan Teknologi Kecerdasan Buatan. Majalah Semi Ilmiah Populer Komunikasi Massa. Kementerian Komunikasi dan Informatika. Vol.2 No.2. Jakarta. Desember 2021

Matthew U. Scherer. ‘Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies’. Harvard Journal of Law & Technology. Center for Democracy and Technology. Vol. 29 No.2. September 2016

O. Strelkova dan O. Pasichnyk. ‘Three Types of Artificial Intelligence’. Khmelnytskyi National University. Ukraina. Mei 2017.

Qur’ani Dewi Kusumawardani. ‘Hukum Progresif dan Perkembangan Teknologi Kecerdasan Buatan’. Veritas et Justitia. Kementerian Komunikasi dan Informatika Republik Indonesia. Vol. 5 No. 1. Jakarta. Juni 2019

Samuel Maireg Biresaw. Abhijit Umesh Saste. The Impacts of Artificial Intelligence on Research in the Legal Profession. International Journal of Law and Society. Science Publishing Group. Vol. 5. No.1. 2022

Steven M. Bellovin. et. al. ‘When enough is enough: Location tracking. mosaic theory. and machine learning’. NYU Journal of Law and Liberty. New York University. Vol. 8(2). New York. 2014

Themistoklis Tzimas. Artificial Intelligence as Global Commons and the “International Law Supremacy” Principle. Advances in Social Science. Education and Humanities Research. Atlantis Press. Vol. 211. 2018

US Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools. National Institute of Science and Technology. Amerika Serikat. 2019.

Vincent Boulanin & Maaike Vebruggen. ‘Mapping The Development of Autonomy in Weapon Systems’. SIPRI. Stockholm. November 2017.

WIPO. Technology Trends Report 2019—Artificial Intelligence. WIPO (2019)

Kamus:

The Oxford English Dictionary. ‘Artificial intelligence’. https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095426960

Berita:

Kanagawa Police to Launch AI-Based Predictive Policing System Before Olympics. The Japan Times. Diakses melalui https://www.japantimes.co.jp/news/2018/01/29/national/crime-legal/kanagawa-police-launch-ai-based-predictive-policing-system-olympics/#.XrSAPqhKiUk

Anonim. Editorial: AI Systems Hold Great Promise for Local gov’ts, but Efficiency isn’t Everything. The Mainichi. (Aug 13 2019). diakses melalui https://mainichi.jp/english/articles/20190813/p2a/00m/0na/012000c

‘Synced. One Year Countdown: Readying AI Security for the Tokyo 2020 Olympics’. https://syncedreview.com/2019/07/20/one-year-countdown-readying-ai-sec urity-for-the-tokyo-2020-olympics/

Skripsi & Tesis:

Jaoa Paulo De Almeida Lenardon. 2017. "*The regulation of Artificial Intelligence"*. Tesis. Master Hukum. Tilburg Institute for Law, Technology, and Society. Tilburg University. Tilburg.

Marcella Sutanto. 2021. *"Perlindungan Hukum Atas Ciptaan Yang Dihasilkan Oleh Kecerdasan Buatan".* Skripsi. Sarjana Hukum. Fakultas Hukum. Universitas Hasanuddin. Makassar.

Laman Internet:

Britannica. 'Artificial Intelligence'. Diakses melalui https://www.britannica.com/technology/artificial-intelligence

----------.'Easter Island'. Diakses melalui https://www.britannica.com/place/Easter-Island

----------.'International Business Machines (IBM)' Diakses melalui https://www.britannica.com/topic/International-Business-Machines-Corporation

Alex Davies. 'Everyone Wants a Level 5 Self-Driving Car—Here's What That Means'. Diakses melalui www.wired.com/2016/08/self-driving-car-levels-sae-nhtsa/

Anonim. 'AI Update: White House Issues 10 Principles for Artificial Intelligence Regulation'. Diakses melalui https://www.insidetechmedia.com/2020/01/14/ai-update-white-house-issues-10-principles-for-artificial-intelligence-regulation/

----------. 'Pengenalan Deep Learning'. Diakses melalui https://machinelearning.mipa.ugm.ac.id/2018/06/10/pengenalan-deep-learning/

----------. 'Speech Recognition'. Diakses melalui https://mti.binus.ac.id/2019/05/08/speech-recognition/

----------. 'Tesla Autopilot – The Ultimate Guide'. Diakses melalui https://www.findmyelectric.com/tesla-autopilot-ultimate-guide/

----------. 'Types of Artificial Intelligence and Examples'. Diakses melalui https://medium.com/predict/types-of-artificial-intelligence-and-examples-4f586489c5de

----------. 'Voice Recognition'. Diakses melalui https://www.computerhope.com/jargon/v/voicreco.htm

----------. 'What Are The Three Types of AI'. Diakses melalui https://www.deccanherald.com/brandspot/pr-spot/what-are-the-3-types-of-ai-853275.html

Ben Lutkevich. ‘Artificial General Intelligence’. Diakses melalui https://searchenterpriseai.techtarget.com/definition/artificial-general-intelligence-AGI

----------. ‘Natural Language Processing’. Diakses melalui https://www.techtarget.com/searchenterpriseai/definition/natural-language-processing-NLP

Bernard Marr. ‘What are the four types of AI’. Diakses melalui https://bernardmarr.com/what-are-the-four-types-of-ai/

Ed Darack. ‘The Drone that Stalked Bin Laden’. Diakses melalui https://www.smithsonianmag.com/air-space-magazine/drone-staked-out-bin-ladens-neighborhood-180958482/

Emily Reynolos. ‘The agony of Sophia. the world’s first robot citizen condemned to a lifeless career in marketing’. Diakses melalui https://www.wired.co.uk/article/sophia-robot-citizen-womens-rights-detriot-become-human-hanson-robotics

Estelle Masse. ‘Data Protection: why it matters and how to protect it’. Diakses melalui https://www.accessnow.org/data-protection-matters-protect/

Europe Union. ‘Call for a High-Level Expert Group on Artificial Intelligence’. Diakses melalui https://ec.europa.eu/digital-single-market/en/news/call-high-level-expert-group-artificial-intelligence

----------. ‘Europe fit for the Digital Age: Commision proposes new rules and actions for excellence and trust in Artificial Intelligence’. Diakses melalui https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682

Faiz Siddiqui et.al. ‘Tesla’s running Autopilot Involved in 273 Crashes Reported Since Last Year’. diakses melalui https://www.washingtonpost.com/technology/2022/06/15/tesla-autopilot-crashes/

Fajar Pebrianto. ‘Otorita Sebut Mobil Tanpa Awak Jadi Tulang Punggung Transportasi IKN’, Diakses melalui https://nasional.tempo.co/read/1593504/otorita-sebut-mobil-tanpa-awak-jadi-tulang-punggung-transportasi-ikn

Gina Taranto. The Evolution of TAR. diakses melalui https://www.law.com/2020/12/31/the-evolution-of-tar/?slreturn=20220621052313

GPAI, ‘The Global Partner on Artificial Intelligence (GPAI)’. Diakses melalui https://gpai.ai/

Graham Webster et.al. ‘China’s New Generation Artificial Intelligence Development Plan’. Diakses melalui https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

IBM. ‘What is computer vision?’. Diakses melalui https://www.ibm.com/topics/computer-vision

-----. ‘Introducing IBM AI Governance’ Diakses melalui https://www.ibm.com/cloud/blog/announcements/introducing-ibm-ai-governance

-----. ‘AI Ethics’ Diakses melalui https://www.ibm.com/artificial-intelligence/ethics

ILO, ‘Skills strategies for future labour markets’. Diakses melalui https://www.ilo.org/skills/areas/skills-training-for-poverty-reduction/lang--en/index.htm

Ivan Mehta. ‘Deepmind’s AlphaZero AI is the new champion in chess. shogi. and Go’. Diakses melalui https://thenextweb.com/artificial-intelligence/2018/12/07/deepminds-alphazero-ai-is-the-new-champion-in-chess-shogi-and-go/

Joshua P. Meltzer. The impact of artificial intelligence on international trade. Diakses melalui https://www.brookings.edu/research/the-impact-of-artificial-intelligence-on-international-trade/

Kevin Roose. ‘An A.I. Generated Picture Won an Art Prize. Artists Aren’t Happy.’ Diakses melalui https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html

Liz Kwo. Contributed: Top 10 Use Cases AI Healthcare. Diakses melalui https://www.mobihealthnews.com/news/contributed-top-10-use-cases-ai-healthcare

Mark Robert Anderson. ‘After 75 years Isaac Asimov’s three laws of robotics need updating’. Diakses melalui https://theconversation.com/after-75-years-isaac-asimovs-three-laws-of-robotics-need-updating-74501

Naveen Joshi. ‘7 Types of Artificial Intelligence’. 19 Juni 2019. Diakses melalui https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence/?sh=7f0a8e77233e

Next Rembrandt. ‘Next Rembrandt’ Diakses melalui https://www.nextrembrandt.com/

Nitasha Tiku. ‘The Google engineer who thinks the company’s AI has come to life’. Diakses melalui https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/

Oracle. What is IoT?. Diakses melalui https://www.oracle.com/internet-of-things/what-is-iot/

‘Principles for Digital Development.’ Diakses melalui https://digitalprinciples.org/about/

Robert F. Trager & Laura M. Luca. ‘Killer Robots Lethal Autonomous Weapons Systems Ukraine-Libya Regulation’. Diakses melalui https://foreignpolicy.com/2022/05/11/killer-robots-lethal-autonomous-weapons-systems-ukraine-libya-regulation/

ROSS. ‘ROSS Intelligence’. Diakses melalui https://www.rossintelligence.com/what-is-ai

Sarah Griffith. ‘This AI Software can tell if you’re at risk from cancer before symptoms appear’. Diakses melalui https://www.wired.co.uk/article/cancer-risk-ai-mammograms

Stanford. Robotics: a brief history. Diakses melalui https://cs.stanford.edu/people/eroberts/courses/soco/projects/1998-99/robotics/history.html

The Government of Japan. ‘How Japan Uses AI and Robotics to Solve Social Issues and Achieve Economic Growth’. Harvard Business Review. Diakses melalui https://hbr.org/sponsored/2020/02/how-japan-uses-ai-and-robotics-to-solve-social-issues-and-achieve-economic-growth

Tyler Rogoway. ‘Meet Israel’s Suicide Squad of Self Sacraficing Drones’. diakses melalui https://www.thedrive.com/the-war-zone/4760/meet-israels-suicide-squad-of-self-sacrificing-drones

UNESCO. ‘History of UNESCO’. Diakses melalui https://www.unesco.org/en/history

Zoe Kleinman. ‘Tesla’s Optimus and the problems with humanoid’. Diakses melalui https://www.bbc.com/news/technology-63130363

LAMPIRAN

EUROPEAN
COMMISSION

Brussels, 21.4.2021
COM(2021) 206 final

2021/0106 (COD)

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

{SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}

EXPLANATORY MEMORANDUM

1. CONTEXT OF THE PROPOSAL

1.1. Reasons for and objectives of the proposal

This explanatory memorandum accompanies the proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Artificial Intelligence (AI) is a fast evolving family of technologies that can bring a wide array of economic and societal benefits across the entire spectrum of industries and social activities. By improving prediction, optimising operations and resource allocation, and personalising service delivery, the use of artificial intelligence can support socially and environmentally beneficial outcomes and provide key competitive advantages to companies and the European economy. Such action is especially needed in high-impact sectors, including climate change, environment and health, the public sector, finance, mobility, home affairs and agriculture. However, the same elements and techniques that power the socio-economic benefits of AI can also bring about new risks or negative consequences for individuals or the society. In light of the speed of technological change and possible challenges, the EU is committed to strive for a balanced approach. It is in the Union interest to preserve the EU's technological leadership and to ensure that Europeans can benefit from new technologies developed and functioning according to Union values, fundamental rights and principles.

This proposal delivers on the political commitment by President von der Leyen, who announced in her political guidelines for the 2019-2024 Commission "A Union that strives for more"[1], that the Commission would put forward legislation for a coordinated European approach on the human and ethical implications of AI. Following on that announcement, on 19 February 2020 the Commission published the White Paper on AI - A European approach to excellence and trust[2]. The White Paper sets out policy options on how to achieve the twin objective of promoting the uptake of AI and of addressing the risks associated with certain uses of such technology. This proposal aims to implement the second objective for the development of an ecosystem of trust by proposing a legal framework for trustworthy AI. The proposal is based on EU values and fundamental rights and aims to give people and other users the confidence to embrace AI-based solutions, while encouraging businesses to develop them. AI should be a tool for people and be a force for good in society with the ultimate aim of increasing human well-being. Rules for AI available in the Union market or otherwise affecting people in the Union should therefore be human centric, so that people can trust that the technology is used in a way that is safe and compliant with the law, including the respect of fundamental rights. Following the publication of the White Paper, the Commission launched a broad stakeholder consultation, which was met with a great interest by a large number of stakeholders who were largely supportive of regulatory intervention to address the challenges and concerns raised by the increasing use of AI.

The proposal also responds to explicit requests from the European Parliament (EP) and the European Council, which have repeatedly expressed calls for legislative action to ensure a well-functioning internal market for artificial intelligence systems ('AI systems') where both benefits and risks of AI are adequately addressed at Union level. It supports the objective of the Union being a global leader in the development of secure, trustworthy and ethical artificial

[1] https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf

[2] European Commission, White Paper on Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65 final, 2020.

intelligence as stated by the European Council[3] and ensures the protection of ethical principles as specifically requested by the European Parliament[4].

In 2017, the European Council called for a 'sense of urgency to address emerging trends' including 'issues such as artificial intelligence …, while at the same time ensuring a high level of data protection, digital rights and ethical standards'[5]. In its 2019 Conclusions on the Coordinated Plan on the development and use of artificial intelligence Made in Europe[6], the Council further highlighted the importance of ensuring that European citizens' rights are fully respected and called for a review of the existing relevant legislation to make it fit for purpose for the new opportunities and challenges raised by AI. The European Council has also called for a clear determination of the AI applications that should be considered high-risk[7].

The most recent Conclusions from 21 October 2020 further called for addressing the opacity, complexity, bias, a certain degree of unpredictability and partially autonomous behaviour of certain AI systems, to ensure their compatibility with fundamental rights and to facilitate the enforcement of legal rules[8].

The European Parliament has also undertaken a considerable amount of work in the area of AI. In October 2020, it adopted a number of resolutions related to AI, including on ethics[9], liability[10] and copyright[11]. In 2021, those were followed by resolutions on AI in criminal matters[12] and in education, culture and the audio-visual sector[13]. The EP Resolution on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies specifically recommends to the Commission to propose legislative action to harness the opportunities and benefits of AI, but also to ensure protection of ethical principles. The resolution includes a text of the legislative proposal for a regulation on ethical principles for the development, deployment and use of AI, robotics and related technologies. In accordance with the political commitment made by President von der Leyen in her Political Guidelines as regards resolutions adopted by the European Parliament under Article 225 TFEU, this

---

3 European Council, Special meeting of the European Council (1 and 2 October 2020) – Conclusions, EUCO 13/20, 2020, p. 6.

4 European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies, 2020/2012(INL).

5 European Council, *European Council meeting (19 October 2017) – Conclusion* EUCO 14/17, 2017, p. 8.

6 Council of the European Union, *Artificial intelligence b) Conclusions on the coordinated plan on artificial intelligence-Adoption* 6177/19, 2019.

7 European Council, *Special meeting of the European Council (1and 2 October 2020) – Conclusions* EUCO 13/20, 2020.

8 Council of the European Union, *Presidency conclusions - The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change*, 11481/20, 2020.

9 European Parliament resolution of 20 October 2020 on a framework of ethical aspects of artificial intelligence, robotics and related technologies, 2020/2012(INL).

10 European Parliament resolution of 20 October 2020 on a civil liability regime for artificial intelligence, 2020/2014(INL).

11 European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies, 2020/2015(INI).

12 European Parliament Draft Report, Artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters, 2020/2016(INI).

13 European Parliament Draft Report, Artificial intelligence in education, culture and the audiovisual sector, 2020/2017(INI). In that regard, the Commission has adopted the Digital Education Action Plan 2021-2027: Resetting education and training for the digital age, which foresees the development of ethical guidelines in AI and Data usage in education – Commission Communication COM(2020) 624 final.

proposal takes into account the aforementioned resolution of the European Parliament in full respect of proportionality, subsidiarity and better law making principles.

Against this political context, the Commission puts forward the proposed regulatory framework on Artificial Intelligence with the following **specific objectives**:

- ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values;
- ensure legal certainty to facilitate investment and innovation in AI;
- enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems;
- facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.

To achieve those objectives, this proposal presents a balanced and proportionate horizontal regulatory approach to AI that is limited to the minimum necessary requirements to address the risks and problems linked to AI, without unduly constraining or hindering technological development or otherwise disproportionately increasing the cost of placing AI solutions on the market. The proposal sets a robust and flexible legal framework. On the one hand, it is comprehensive and future-proof in its fundamental regulatory choices, including the principle-based requirements that AI systems should comply with. On the other hand, it puts in place a proportionate regulatory system centred on a well-defined risk-based regulatory approach that does not create unnecessary restrictions to trade, whereby legal intervention is tailored to those concrete situations where there is a justified cause for concern or where such concern can reasonably be anticipated in the near future. At the same time, the legal framework includes flexible mechanisms that enable it to be dynamically adapted as the technology evolves and new concerning situations emerge.

The proposal sets harmonised rules for the development, placement on the market and use of AI systems in the Union following a proportionate risk-based approach. It proposes a single future-proof definition of AI. Certain particularly harmful AI practices are prohibited as contravening Union values, while specific restrictions and safeguards are proposed in relation to certain uses of remote biometric identification systems for the purpose of law enforcement. The proposal lays down a solid risk methodology to define "high-risk" AI systems that pose significant risks to the health and safety or fundamental rights of persons. Those AI systems will have to comply with a set of horizontal mandatory requirements for trustworthy AI and follow conformity assessment procedures before those systems can be placed on the Union market. Predictable, proportionate and clear obligations are also placed on providers and users of those systems to ensure safety and respect of existing legislation protecting fundamental rights throughout the whole AI systems' lifecycle. For some specific AI systems, only minimum transparency obligations are proposed, in particular when chatbots or 'deep fakes' are used.

The proposed rules will be enforced through a governance system at Member States level, building on already existing structures, and a cooperation mechanism at Union level with the establishment of a European Artificial Intelligence Board. Additional measures are also proposed to support innovation, in particular through AI regulatory sandboxes and other measures to reduce the regulatory burden and to support Small and Medium-Sized Enterprises ('SMEs') and start-ups.

1.2. Consistency with existing policy provisions in the policy area

The horizontal nature of the proposal requires full consistency with existing Union legislation applicable to sectors where high-risk AI systems are already used or likely to be used in the near future.

Consistency is also ensured with the EU Charter of Fundamental Rights and the existing secondary Union legislation on data protection, consumer protection, non-discrimination and gender equality. The proposal is without prejudice and complements the General Data Protection Regulation (Regulation (EU) 2016/679) and the Law Enforcement Directive (Directive (EU) 2016/680) with a set of harmonised rules applicable to the design, development and use of certain high-risk AI systems and restrictions on certain uses of remote biometric identification systems. Furthermore, the proposal complements existing Union law on non-discrimination with specific requirements that aim to minimise the risk of algorithmic discrimination, in particular in relation to the design and the quality of data sets used for the development of AI systems complemented with obligations for testing, risk management, documentation and human oversight throughout the AI systems' lifecycle. The proposal is without prejudice to the application of Union competition law.

As regards high-risk AI systems which are safety components of products, this proposal will be integrated into the existing sectoral safety legislation to ensure consistency, avoid duplications and minimise additional burdens. In particular, as regards high-risk AI systems related to products covered by the New Legislative Framework (NLF) legislation (e.g. machinery, medical devices, toys), the requirements for AI systems set out in this proposal will be checked as part of the existing conformity assessment procedures under the relevant NLF legislation. With regard to the interplay of requirements, while the safety risks specific to AI systems are meant to be covered by the requirements of this proposal, NLF legislation aims at ensuring the overall safety of the final product and therefore may contain specific requirements regarding the safe integration of an AI system into the final product. The proposal for a Machinery Regulation, which is adopted on the same day as this proposal fully reflects this approach. As regards high-risk AI systems related to products covered by relevant Old Approach legislation (e.g. aviation, cars), this proposal would not directly apply. However, the ex-ante essential requirements for high-risk AI systems set out in this proposal will have to be taken into account when adopting relevant implementing or delegated legislation under those acts.

As regards AI systems provided or used by regulated credit institutions, the authorities responsible for the supervision of the Union's financial services legislation should be designated as competent authorities for supervising the requirements in this proposal to ensure a coherent enforcement of the obligations under this proposal and the Union's financial services legislation where AI systems are to some extent implicitly regulated in relation to the internal governance system of credit institutions. To further enhance consistency, the conformity assessment procedure and some of the providers' procedural obligations under this proposal are integrated into the procedures under Directive 2013/36/EU on access to the activity of credit institutions and the prudential supervision[14].

[14] Directive 2013/36/EU of the European Parliament and of the Council of 26 June 2013 on access to the activity of credit institutions and the prudential supervision of credit institutions and investment firms, amending Directive 2002/87/EC and repealing Directives 2006/48/EC and 2006/49/EC Text with EEA relevance, OJ L 176, 27.6.2013, p. 338–436.

This proposal is also consistent with the applicable Union legislation on services, including on intermediary services regulated by the e-Commerce Directive 2000/31/EC[15] and the Commission's recent proposal for the Digital Services Act (DSA)[16].

In relation to AI systems that are components of large-scale IT systems in the Area of Freedom, Security and Justice managed by the European Union Agency for the Operational Management of Large-Scale IT Systems (eu-LISA), the proposal will not apply to those AI systems that have been placed on the market or put into service before one year has elapsed from the date of application of this Regulation, unless the replacement or amendment of those legal acts leads to a significant change in the design or intended purpose of the AI system or AI systems concerned.

1.3. Consistency with other Union policies

The proposal is part of a wider comprehensive package of measures that address problems posed by the development and use of AI, as examined in the White Paper on AI. Consistency and complementarity is therefore ensured with other ongoing or planned initiatives of the Commission that also aim to address those problems, including the revision of sectoral product legislation (e.g. the Machinery Directive, the General Product Safety Directive) and initiatives that address liability issues related to new technologies, including AI systems. Those initiatives will build on and complement this proposal in order to bring legal clarity and foster the development of an ecosystem of trust in AI in Europe.

The proposal is also coherent with the Commission's overall digital strategy in its contribution to promoting technology that works for people, one of the three main pillars of the policy orientation and objectives announced in the Communication 'Shaping Europe's digital future'[17]. It lays down a coherent, effective and proportionate framework to ensure AI is developed in ways that respect people's rights and earn their trust, making Europe fit for the digital age and turning the next ten years into the **Digital Decade**[18].

Furthermore, the promotion of AI-driven innovation is closely linked to the **Data Governance Act**[19]**,** the **Open Data Directive**[20] and other initiatives under **the EU strategy for data**[21], which will establish trusted mechanisms and services for the re-use, sharing and pooling of data that are essential for the development of data-driven AI models of high quality.

The proposal also strengthens significantly the Union's role to help shape global norms and standards and promote trustworthy AI that is consistent with Union values and interests. It provides the Union with a powerful basis to engage further with its external partners, including third countries, and at international fora on issues relating to AI.

---

15 Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'), OJ L 178, 17.7.2000, p. 1–16.

16 See Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC COM/2020/825 final.

17 Communication from the Commission, Shaping Europe's Digital Future, COM/2020/67 final.

18 2030 Digital Compass: the European way for the Digital Decade.

19 Proposal for a Regulation on European data governance (Data Governance Act) COM/2020/767.

20 Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, PE/28/2019/REV/1, OJ L 172, 26.6.2019, p. 56–83.

21 Commission Communication, A European strategy for data COM/2020/66 final.

2. LEGAL BASIS, SUBSIDIARITY AND PROPORTIONALITY

2.1. Legal basis

The legal basis for the proposal is in the first place Article 114 of the Treaty on the Functioning of the European Union (TFEU), which provides for the adoption of measures to ensure the establishment and functioning of the internal market.

This proposal constitutes a core part of the EU digital single market strategy. The primary objective of this proposal is to ensure the proper functioning of the internal market by setting harmonised rules in particular on the development, placing on the Union market and the use of products and services making use of AI technologies or provided as stand-alone AI systems. Some Member States are already considering national rules to ensure that AI is safe and is developed and used in compliance with fundamental rights obligations. This will likely lead to two main problems: i) a fragmentation of the internal market on essential elements regarding in particular the requirements for the AI products and services, their marketing, their use, the liability and the supervision by public authorities, and ii) the substantial diminishment of legal certainty for both providers and users of AI systems on how existing and new rules will apply to those systems in the Union. Given the wide circulation of products and services across borders, these two problems can be best solved through EU harmonizing legislation.

Indeed, the proposal defines common mandatory requirements applicable to the design and development of certain AI systems before they are placed on the market that will be further operationalised through harmonised technical standards. The proposal also addresses the situation after AI systems have been placed on the market by harmonising the way in which ex-post controls are conducted.

In addition, considering that this proposal contains certain specific rules on the protection of individuals with regard to the processing of personal data, notably restrictions of the use of AI systems for 'real-time' remote biometric identification in publicly accessible spaces for the purpose of law enforcement, it is appropriate to base this regulation, in as far as those specific rules are concerned, on Article 16 of the TFEU.

2.2. Subsidiarity (for non-exclusive competence)

The nature of AI, which often relies on large and varied datasets and which may be embedded in any product or service circulating freely within the internal market, entails that the objectives of this proposal cannot be effectively achieved by Member States alone. Furthermore, an emerging patchwork of potentially divergent national rules will hamper the seamless circulation of products and services related to AI systems across the EU and will be ineffective in ensuring the safety and protection of fundamental rights and Union values across the different Member States. National approaches in addressing the problems will only create additional legal uncertainty and barriers, and will slow market uptake of AI.

The objectives of this proposal can be better achieved at Union level to avoid a further fragmentation of the Single Market into potentially contradictory national frameworks preventing the free circulation of goods and services embedding AI. A solid European regulatory framework for trustworthy AI will also ensure a level playing field and protect all people, while strengthening Europe's competitiveness and industrial basis in AI. Only common action at Union level can also protect the Union's digital sovereignty and leverage its tools and regulatory powers to shape global rules and standards.

2.3. Proportionality

The proposal builds on existing legal frameworks and is proportionate and necessary to achieve its objectives, since it follows a risk-based approach and imposes regulatory burdens only when an AI system is likely to pose high risks to fundamental rights and safety. For other, non-high-risk AI systems, only very limited transparency obligations are imposed, for example in terms of the provision of information to flag the use of an AI system when interacting with humans. For high-risk AI systems, the requirements of high quality data, documentation and traceability, transparency, human oversight, accuracy and robustness, are strictly necessary to mitigate the risks to fundamental rights and safety posed by AI and that are not covered by other existing legal frameworks. Harmonised standards and supporting guidance and compliance tools will assist providers and users in complying with the requirements laid down by the proposal and minimise their costs. The costs incurred by operators are proportionate to the objectives achieved and the economic and reputational benefits that operators can expect from this proposal.

2.4. Choice of the instrument

The choice of a regulation as a legal instrument is justified by the need for a uniform application of the new rules, such as definition of AI, the prohibition of certain harmful AI-enabled practices and the classification of certain AI systems. The direct applicability of a Regulation, in accordance with Article 288 TFEU, will reduce legal fragmentation and facilitate the development of a single market for lawful, safe and trustworthy AI systems. It will do so, in particular, by introducing a harmonised set of core requirements with regard to AI systems classified as high-risk and obligations for providers and users of those systems, improving the protection of fundamental rights and providing legal certainty for operators and consumers alike.

At the same time, the provisions of the regulation are not overly prescriptive and leave room for different levels of Member State action for elements that do not undermine the objectives of the initiative, in particular the internal organisation of the market surveillance system and the uptake of measures to foster innovation.

3. RESULTS OF EX-POST EVALUATIONS, STAKEHOLDER CONSULTATIONS AND IMPACT ASSESSMENTS

3.1. Stakeholder consultation

This proposal is the result of extensive consultation with all major stakeholders, in which the general principles and minimum standards for consultation of interested parties by the Commission were applied.

An **online public consultation** was launched on 19 February 2020 along with the publication of the White Paper on Artificial Intelligence and ran until 14 June 2020. The objective of that consultation was to collect views and opinions on the White Paper. It targeted all interested stakeholders from the public and private sectors, including governments, local authorities, commercial and non-commercial organisations, social partners, experts, academics and citizens. After analysing all the responses received, the Commission published a summary outcome and the individual responses on its website[22].

In total, 1215 contributions were received, of which 352 were from companies or business organisations/associations, 406 from individuals (92%individuals from EU), 152 on behalf of

[22] See all consultation results here.

academic/research institutions, and 73 from public authorities. Civil society's voices were represented by 160 respondents (among which 9 consumers' organisations, 129 non-governmental organisations and 22 trade unions), 72 respondents contributed as 'others'. Of the 352 business and industry representatives, 222 were companies and business representatives, 41.5% of which were micro, small and medium-sized enterprises. The rest were business associations. Overall, 84% of business and industry replies came from the EU-27. Depending on the question, between 81 and 598 of the respondents used the free text option to insert comments. Over 450 position papers were submitted through the EU Survey website, either in addition to questionnaire answers (over 400) or as stand-alone contributions (over 50).

Overall, there is a general agreement amongst stakeholders on a need for action. A large majority of stakeholders agree that legislative gaps exist or that new legislation is needed. However, several stakeholders warn the Commission to avoid duplication, conflicting obligations and overregulation. There were many comments underlining the importance of a technology neutral and proportionate regulatory framework.

Stakeholders mostly requested a narrow, clear and precise definition for AI. Stakeholders also highlighted that besides the clarification of the term of AI, it is important to define 'risk', 'high-risk', 'low-risk', 'remote biometric identification' and 'harm'.

Most of the respondents are explicitly in favour of the risk-based approach. Using a risk-based framework was considered a better option than blanket regulation of all AI systems. The types of risks and threats should be based on a sector-by-sector and case-by-case approach. Risks also should be calculated taking into account the impact on rights and safety.

Regulatory sandboxes could be very useful for the promotion of AI and are welcomed by certain stakeholders, especially the Business Associations.

Among those who formulated their opinion on the enforcement models, more than 50%, especially from the business associations, were in favour of a combination of an ex-ante risk self-assessment and an ex-post enforcement for high-risk AI systems.

3.2. Collection and use of expertise

The proposal builds on two years of analysis and close involvement of stakeholders, including academics, businesses, social partners, non-governmental organisations, Member States and citizens. The preparatory work started in 2018 with the setting up of a **High-Level Expert Group on AI (HLEG)** which had an inclusive and broad composition of 52 well-known experts tasked to advise the Commission on the implementation of the Commission's Strategy on Artificial Intelligence. In April 2019, the Commission supported[23] the key requirements set out in the HLEG ethics guidelines for Trustworthy AI[24], which had been revised to take into account more than 500 submissions from stakeholders. The key requirements reflect a widespread and common approach, as evidenced by a plethora of ethical codes and principles developed by many private and public organisations in Europe and beyond, that AI development and use should be guided by certain essential value-oriented principles. The Assessment List for Trustworthy Artificial Intelligence (ALTAI)[25] made those requirements operational in a piloting process with over 350 organisations.

---

23 European Commission, *Building Trust in Human-Centric Artificial Intelligence*, COM(2019) 168.

24 HLEG, *Ethics Guidelines for Trustworthy AI*, 2019.

25 HLEG, *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*, 2020.

In addition, the **AI Alliance**[26] was formed as a platform for approximately 4000 stakeholders to debate the technological and societal implications of AI, culminating in a yearly AI Assembly.

The **White Paper** on AI further developed this inclusive approach, inciting comments from more than 1250 stakeholders, including over 450 additional position papers. As a result, the Commission published an Inception Impact Assessment, which in turn attracted more than 130 comments[27]. **Additional stakeholder workshops and events** were also organised the results of which support the analysis in the impact assessment and the policy choices made in this proposal[28]. An **external study** was also procured to feed into the impact assessment.

3.3. Impact assessment

In line with its "Better Regulation" policy, the Commission conducted an impact assessment for this proposal examined by the Commission's Regulatory Scrutiny Board. A meeting with the Regulatory Scrutiny Board was held on 16 December 2020, which was followed by a negative opinion. After substantial revision of the impact assessment to address the comments and a resubmission of the impact assessment, the Regulatory Scrutiny Board issued a positive opinion on 21 March 2021. The opinions of the Regulatory Scrutiny Board, the recommendations and an explanation of how they have been taken into account are presented in Annex 1 of the impact assessment.

The Commission examined different policy options to achieve the general objective of the proposal, which is to **ensure the proper functioning of the single market** by creating the conditions for the development and use of trustworthy AI in the Union.

Four policy options of different degrees of regulatory intervention were assessed:

- **Option 1**: EU legislative instrument setting up a voluntary labelling scheme;
- **Option 2**: a sectoral, "ad-hoc" approach;
- **Option 3**: Horizontal EU legislative instrument following a proportionate risk-based approach;
- **Option 3+**: Horizontal EU legislative instrument following a proportionate risk-based approach + codes of conduct for non-high-risk AI systems;
- **Option 4**: Horizontal EU legislative instrument establishing mandatory requirements for all AI systems, irrespective of the risk they pose.

According to the Commission's established methodology, each policy option was evaluated against economic and societal impacts, with a particular focus on impacts on fundamental rights. The preferred option is option 3+, a regulatory framework for high-risk AI systems only, with the possibility for all providers of non-high-risk AI systems to follow a code of conduct. The requirements will concern data, documentation and traceability, provision of information and transparency, human oversight and robustness and accuracy and would be mandatory for high-risk AI systems. Companies that introduced codes of conduct for other AI systems would do so voluntarily.

---

[26] The AI Alliance is a multi-stakeholder forum launched in June 2018, AI Alliance https://ec.europa.eu/digital-single-market/en/european-ai-alliance

[27] European Commission, *Inception Impact Assessment For a Proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence*.

[28] For details of all the consultations that have been carried out see Annex 2 of the impact assessment.

The preferred option was considered suitable to address in the most effective way the objectives of this proposal. By requiring a restricted yet effective set of actions from AI developers and users, the preferred option limits the risks of violation of fundamental rights and safety of people and foster effective supervision and enforcement, by targeting the requirements only to systems where there is a high risk that such violations could occur. As a result, that option keeps compliance costs to a minimum, thus avoiding an unnecessary slowing of uptake due to higher prices and compliance costs. In order to address possible disadvantages for SMEs, this option includes several provisions to support their compliance and reduce their costs, including creation of regulatory sandboxes and obligation to consider SMEs interests when setting fees related to conformity assessment.

The preferred option will increase people's trust in AI, companies will gain in legal certainty, and Member States will see no reason to take unilateral action that could fragment the single market. As a result of higher demand due to higher trust, more available offers due to legal certainty, and the absence of obstacles to cross-border movement of AI systems, the single market for AI will likely flourish. The European Union will continue to develop a fast-growing AI ecosystem of innovative services and products embedding AI technology or stand-alone AI systems, resulting in increased digital autonomy.

Businesses or public authorities that develop or use AI applications that constitute a high risk for the safety or fundamental rights of citizens would have to comply with specific requirements and obligations. Compliance with these requirements would imply costs amounting to approximately EUR € 6000 to EUR € 7000 for the supply of an average high-risk AI system of around EUR € 170000 by 2025. For AI users, there would also be the annual cost for the time spent on ensuring human oversight where this is appropriate, depending on the use case. Those have been estimated at approximately EUR € 5000 to EUR € 8000 per year. Verification costs could amount to another EUR € 3000 to EUR € 7500 for suppliers of high-risk AI. Businesses or public authorities that develop or use any AI applications not classified as high risk would only have minimal obligations of information. However, they could choose to join others and together adopt a code of conduct to follow suitable requirements, and to ensure that their AI systems are trustworthy. In such a case, costs would be at most as high as for high-risk AI systems, but most probably lower.

The impacts of the policy options on different categories of stakeholders (economic operators/ business; conformity assessment bodies, standardisation bodies and other public bodies; individuals/citizens; researchers) are explained in detail in Annex 3 of the Impact assessment supporting this proposal.

3.4. Regulatory fitness and simplification

This proposal lays down obligation that will apply to providers and users of high-risk AI systems. For providers who develop and place such systems on the Union market, it will create legal certainty and ensure that no obstacle to the cross-border provision of AI-related services and products emerge. For companies using AI, it will promote trust among their customers. For national public administrations, it will promote public trust in the use of AI and strengthen enforcement mechanisms (by introducing a European coordination mechanism, providing for appropriate capacities, and facilitating audits of the AI systems with new requirements for documentation, traceability and transparency). Moreover, the framework will envisage specific measures supporting innovation, including regulatory sandboxes and specific measures supporting small-scale users and providers of high-risk AI systems to comply with the new rules.

The proposal also specifically aims at strengthening Europe's competitiveness and industrial basis in AI. Full consistency is ensured with existing sectoral Union legislation applicable to

AI systems (e.g. on products and services) that will bring further clarity and simplify the enforcement of the new rules.

3.5. Fundamental rights

The use of AI with its specific characteristics (e.g. opacity, complexity, dependency on data, autonomous behaviour) can adversely affect a number of fundamental rights enshrined in the EU Charter of Fundamental Rights ('the Charter'). This proposal seeks to ensure a high level of protection for those fundamental rights and aims to address various sources of risks through a clearly defined risk-based approach. With a set of requirements for trustworthy AI and proportionate obligations on all value chain participants, the proposal will enhance and promote the protection of the rights protected by the Charter: the right to human dignity (Article 1), respect for private life and protection of personal data (Articles 7 and 8), non-discrimination (Article 21) and equality between women and men (Article 23). It aims to prevent a chilling effect on the rights to freedom of expression (Article 11) and freedom of assembly (Article 12), to ensure protection of the right to an effective remedy and to a fair trial, the rights of defence and the presumption of innocence (Articles 47 and 48), as well as the general principle of good administration. Furthermore, as applicable in certain domains, the proposal will positively affect the rights of a number of special groups, such as the workers' rights to fair and just working conditions (Article 31), a high level of consumer protection (Article 28), the rights of the child (Article 24) and the integration of persons with disabilities (Article 26). The right to a high level of environmental protection and the improvement of the quality of the environment (Article 37) is also relevant, including in relation to the health and safety of people. The obligations for ex ante testing, risk management and human oversight will also facilitate the respect of other fundamental rights by minimising the risk of erroneous or biased AI-assisted decisions in critical areas such as education and training, employment, important services, law enforcement and the judiciary. In case infringements of fundamental rights still happen, effective redress for affected persons will be made possible by ensuring transparency and traceability of the AI systems coupled with strong ex post controls.

This proposal imposes some restrictions on the freedom to conduct business (Article 16) and the freedom of art and science (Article 13) to ensure compliance with overriding reasons of public interest such as health, safety, consumer protection and the protection of other fundamental rights ('responsible innovation') when high-risk AI technology is developed and used. Those restrictions are proportionate and limited to the minimum necessary to prevent and mitigate serious safety risks and likely infringements of fundamental rights.

The increased transparency obligations will also not disproportionately affect the right to protection of intellectual property (Article 17(2)), since they will be limited only to the minimum necessary information for individuals to exercise their right to an effective remedy and to the necessary transparency towards supervision and enforcement authorities, in line with their mandates. Any disclosure of information will be carried out in compliance with relevant legislation in the field, including Directive 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure. When public authorities and notified bodies need to be given access to confidential information or source code to examine compliance with substantial obligations, they are placed under binding confidentiality obligations.

4. BUDGETARY IMPLICATIONS

Member States will have to designate supervisory authorities in charge of implementing the legislative requirements. Their supervisory function could build on existing arrangements, for

example regarding conformity assessment bodies or market surveillance, but would require sufficient technological expertise and human and financial resources. Depending on the pre-existing structure in each Member State, this could amount to 1 to 25 Full Time Equivalents per Member State.

A detailed overview of the costs involved is provided in the 'financial statement' linked to this proposal.

5. OTHER ELEMENTS

5.1. Implementation plans and monitoring, evaluation and reporting arrangements

Providing for a robust monitoring and evaluation mechanism is crucial to ensure that the proposal will be effective in achieving its specific objectives. The Commission will be in charge of monitoring the effects of the proposal. It will establish a system for registering stand-alone high-risk AI applications in a public EU-wide database. This registration will also enable competent authorities, users and other interested people to verify if the high-risk AI system complies with the requirements laid down in the proposal and to exercise enhanced oversight over those AI systems posing high risks to fundamental rights. To feed this database, AI providers will be obliged to provide meaningful information about their systems and the conformity assessment carried out on those systems.

Moreover, AI providers will be obliged to inform national competent authorities about serious incidents or malfunctioning that constitute a breach of fundamental rights obligations as soon as they become aware of them, as well as any recalls or withdrawals of AI systems from the market. National competent authorities will then investigate the incidents/or malfunctioning, collect all the necessary information and regularly transmit it to the Commission with adequate metadata. The Commission will complement this information on the incidents by a comprehensive analysis of the overall market for AI.

The Commission will publish a report evaluating and reviewing the proposed AI framework five years following the date on which it becomes applicable.

5.2. Detailed explanation of the specific provisions of the proposal

5.2.1. SCOPE AND DEFINITIONS (TITLE I)

Title I defines the subject matter of the regulation and the scope of application of the new rules that cover the placing on the market, putting into service and use of AI systems. It also sets out the definitions used throughout the instrument. The definition of AI system in the legal framework aims to be as technology neutral and future proof as possible, taking into account the fast technological and market developments related to AI. In order to provide the needed legal certainty, Title I is complemented by Annex I, which contains a detailed list of approaches and techniques for the development of AI to be adapted by the Commission in line with new technological developments. Key participants across the AI value chain are also clearly defined such as providers and users of AI systems that cover both public and private operators to ensure a level playing field.

5.2.2. PROHIBITED ARTIFICIAL INTELLIGENCE PRACTICES (TITLE II)

Title II establishes a list of prohibited AI. The regulation follows a risk-based approach, differentiating between uses of AI that create (i) an unacceptable risk, (ii) a high risk, and (iii) low or minimal risk. The list of prohibited practices in Title II comprises all those AI systems whose use is considered unacceptable as contravening Union values, for instance by violating fundamental rights. The prohibitions covers practices that have a significant potential to manipulate persons through subliminal techniques beyond their consciousness or exploit

vulnerabilities of specific vulnerable groups such as children or persons with disabilities in order to materially distort their behaviour in a manner that is likely to cause them or another person psychological or physical harm. Other manipulative or exploitative practices affecting adults that might be facilitated by AI systems could be covered by the existing data protection, consumer protection and digital service legislation that guarantee that natural persons are properly informed and have free choice not to be subject to profiling or other practices that might affect their behaviour. The proposal also prohibits AI-based social scoring for general purposes done by public authorities. Finally, the use of 'real time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement is also prohibited unless certain limited exceptions apply.

5.2.3. HIGH-RISK AI SYSTEMS (TITLE III)

Title III contains specific rules for AI systems that create a high risk to the health and safety or fundamental rights of natural persons. In line with a risk-based approach, those high-risk AI systems are permitted on the European market subject to compliance with certain mandatory requirements and an ex-ante conformity assessment. The classification of an AI system as high-risk is based on the intended purpose of the AI system, in line with existing product safety legislation. Therefore, the classification as high-risk does not only depend on the function performed by the AI system, but also on the specific purpose and modalities for which that system is used.

Chapter 1 of Title III sets the classification rules and identifies two main categories of high-risk AI systems:

- AI systems intended to be used as safety component of products that are subject to third party ex-ante conformity assessment;
- other stand-alone AI systems with mainly fundamental rights implications that are explicitly listed in Annex III.

This list of high-risk AI systems in Annex III contains a limited number of AI systems whose risks have already materialised or are likely to materialise in the near future. To ensure that the regulation can be adjusted to emerging uses and applications of AI, the Commission may expand the list of high-risk AI systems used within certain pre-defined areas, by applying a set of criteria and risk assessment methodology.

Chapter 2 sets out the legal requirements for high-risk AI systems in relation to data and data governance, documentation and recording keeping, transparency and provision of information to users, human oversight, robustness, accuracy and security. The proposed minimum requirements are already state-of-the-art for many diligent operators and the result of two years of preparatory work, derived from the Ethics Guidelines of the HLEG[29], piloted by more than 350 organisations[30]. They are also largely consistent with other international recommendations and principles, which ensures that the proposed AI framework is compatible with those adopted by the EU's international trade partners. The precise technical solutions to achieve compliance with those requirements may be provided by standards or by other technical specifications or otherwise be developed in accordance with general engineering or scientific knowledge at the discretion of the provider of the AI system. This flexibility is particularly important, because it allows providers of AI systems to choose the

[29] High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, 2019.

[30] They were also endorsed by the Commission in its 2019 Communication on human-centric approach to AI.

way to meet their requirements, taking into account the state-of-the-art and technological and scientific progress in this field.

Chapter 3 places a clear set of horizontal obligations on providers of high-risk AI systems. Proportionate obligations are also placed on users and other participants across the AI value chain (e.g., importers, distributors, authorized representatives).

Chapter 4 sets the framework for notified bodies to be involved as independent third parties in conformity assessment procedures, while Chapter 5 explains in detail the conformity assessment procedures to be followed for each type of high-risk AI system. The conformity assessment approach aims to minimise the burden for economic operators as well as for notified bodies, whose capacity needs to be progressively ramped up over time. AI systems intended to be used as safety components of products that are regulated under the New Legislative Framework legislation (e.g. machinery, toys, medical devices, etc.) will be subject to the same ex-ante and ex-post compliance and enforcement mechanisms of the products of which they are a component. The key difference is that the ex-ante and ex-post mechanisms will ensure compliance not only with the requirements established by sectorial legislation, but also with the requirements established by this regulation.

As regards stand-alone high-risk AI systems that are referred to in Annex III, a new compliance and enforcement system will be established. This follows the model of the New Legislative Framework legislation implemented through internal control checks by the providers with the exception of remote biometric identification systems that would be subject to third party conformity assessment. A comprehensive ex-ante conformity assessment through internal checks, combined with a strong ex-post enforcement, could be an effective and reasonable solution for those systems, given the early phase of the regulatory intervention and the fact the AI sector is very innovative and expertise for auditing is only now being accumulated. An assessment through internal checks for 'stand-alone' high-risk AI systems would require a full, effective and properly documented ex ante compliance with all requirements of the regulation and compliance with robust quality and risk management systems and post-market monitoring. After the provider has performed the relevant conformity assessment, it should register those stand-alone high-risk AI systems in an EU database that will be managed by the Commission to increase public transparency and oversight and strengthen ex post supervision by competent authorities. By contrast, for reasons of consistency with the existing product safety legislation, the conformity assessments of AI systems that are safety components of products will follow a system with third party conformity assessment procedures already established under the relevant sectoral product safety legislation. New ex ante re-assessments of the conformity will be needed in case of substantial modifications to the AI systems (and notably changes which go beyond what is pre-determined by the provider in its technical documentation and checked at the moment of the ex-ante conformity assessment).

5.2.4. TRANSPARENCY OBLIGATIONS FOR CERTAIN AI SYSTEMS (TITLE IV)

Title IV concerns certain AI systems to take account of the specific risks of manipulation they pose. Transparency obligations will apply for systems that (i) interact with humans, (ii) are used to detect emotions or determine association with (social) categories based on biometric data, or (iii) generate or manipulate content ('deep fakes'). When persons interact with an AI system or their emotions or characteristics are recognised through automated means, people must be informed of that circumstance. If an AI system is used to generate or manipulate image, audio or video content that appreciably resembles authentic content, there should be an obligation to disclose that the content is generated through automated means, subject to

exceptions for legitimate purposes (law enforcement, freedom of expression). This allows persons to make informed choices or step back from a given situation.

5.2.5. MEASURES IN SUPPORT OF INNOVATION (TITLE V)

Title V contributes to the objective to create a legal framework that is innovation-friendly, future-proof and resilient to disruption. To that end, it encourages national competent authorities to set up regulatory sandboxes and sets a basic framework in terms of governance, supervision and liability. AI regulatory sandboxes establish a controlled environment to test innovative technologies for a limited time on the basis of a testing plan agreed with the competent authorities. Title V also contains measures to reduce the regulatory burden on SMEs and start-ups.

5.2.6. GOVERNANCE AND IMPLEMENTATION (TITLES VI, VII AND VII)

Title VI sets up the governance systems at Union and national level. At Union level, the proposal establishes a European Artificial Intelligence Board (the 'Board'), composed of representatives from the Member States and the Commission. The Board will facilitate a smooth, effective and harmonised implementation of this regulation by contributing to the effective cooperation of the national supervisory authorities and the Commission and providing advice and expertise to the Commission. It will also collect and share best practices among the Member States.

At national level, Member States will have to designate one or more national competent authorities and, among them, the national supervisory authority, for the purpose of supervising the application and implementation of the regulation. The European Data Protection Supervisor will act as the competent authority for the supervision of the Union institutions, agencies and bodies when they fall within the scope of this regulation.

Title VII aims to facilitate the monitoring work of the Commission and national authorities through the establishment of an EU-wide database for stand-alone high-risk AI systems with mainly fundamental rights implications. The database will be operated by the Commission and provided with data by the providers of the AI systems, who will be required to register their systems before placing them on the market or otherwise putting them into service.

Title VIII sets out the monitoring and reporting obligations for providers of AI systems with regard to post-market monitoring and reporting and investigating on AI-related incidents and malfunctioning. Market surveillance authorities would also control the market and investigate compliance with the obligations and requirements for all high-risk AI systems already placed on the market. Market surveillance authorities would have all powers under Regulation (EU) 2019/1020 on market surveillance. Ex-post enforcement should ensure that once the AI system has been put on the market, public authorities have the powers and resources to intervene in case AI systems generate unexpected risks, which warrant rapid action. They will also monitor compliance of operators with their relevant obligations under the regulation. The proposal does not foresee the automatic creation of any additional bodies or authorities at Member State level. Member States may therefore appoint (and draw upon the expertise of) existing sectorial authorities, who would be entrusted also with the powers to monitor and enforce the provisions of the regulation.

All this is without prejudice to the existing system and allocation of powers of ex-post enforcement of obligations regarding fundamental rights in the Member States. When necessary for their mandate, existing supervision and enforcement authorities will also have the power to request and access any documentation maintained following this regulation and, where needed, request market surveillance authorities to organise testing of the high-risk AI system through technical means.

5.2.7. CODES OF CONDUCT (TITLE IX)

Title IX creates a framework for the creation of codes of conduct, which aim to encourage providers of non-high-risk AI systems to apply voluntarily the mandatory requirements for high-risk AI systems (as laid out in Title III). Providers of non-high-risk AI systems may create and implement the codes of conduct themselves. Those codes may also include voluntary commitments related, for example, to environmental sustainability, accessibility for persons with disability, stakeholders' participation in the design and development of AI systems, and diversity of development teams.

5.2.8. FINAL PROVISIONS (TITLES X, XI AND XII)

Title X emphasizes the obligation of all parties to respect the confidentiality of information and data and sets out rules for the exchange of information obtained during the implementation of the regulation. Title X also includes measures to ensure the effective implementation of the regulation through effective, proportionate, and dissuasive penalties for infringements of the provisions.

Title XI sets out rules for the exercise of delegation and implementing powers. The proposal empowers the Commission to adopt, where appropriate, implementing acts to ensure uniform application of the regulation or delegated acts to update or complement the lists in Annexes I to VII.

Title XII contains an obligation for the Commission to assess regularly the need for an update of Annex III and to prepare regular reports on the evaluation and review of the regulation. It also lays down final provisions, including a differentiated transitional period for the initial date of the applicability of the regulation to facilitate the smooth implementation for all parties concerned.

2021/0106 (COD)

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

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

Having regard to the Treaty on the Functioning of the European Union, and in particular Articles 16 and 114 thereof,

Having regard to the proposal from the European Commission,

After transmission of the draft legislative act to the national parliaments,

Having regard to the opinion of the European Economic and Social Committee[31],

Having regard to the opinion of the Committee of the Regions[32],

Acting in accordance with the ordinary legislative procedure,

Whereas:

(1) The purpose of this Regulation is to improve the functioning of the internal market by laying down a uniform legal framework in particular for the development, marketing and use of artificial intelligence in conformity with Union values. This Regulation pursues a number of overriding reasons of public interest, such as a high level of protection of health, safety and fundamental rights, and it ensures the free movement of AI-based goods and services cross-border, thus preventing Member States from imposing restrictions on the development, marketing and use of AI systems, unless explicitly authorised by this Regulation.

(2) Artificial intelligence systems (AI systems) can be easily deployed in multiple sectors of the economy and society, including cross border, and circulate throughout the Union. Certain Member States have already explored the adoption of national rules to ensure that artificial intelligence is safe and is developed and used in compliance with fundamental rights obligations. Differing national rules may lead to fragmentation of the internal market and decrease legal certainty for operators that develop or use AI systems. A consistent and high level of protection throughout the Union should therefore be ensured, while divergences hampering the free circulation of AI systems and related products and services within the internal market should be prevented, by laying down uniform obligations for operators and guaranteeing the uniform protection of overriding reasons of public interest and of rights of persons throughout the internal market based on Article 114 of the Treaty on the Functioning of the European Union (TFEU). To the extent that this Regulation contains specific rules on the protection of individuals with regard to the processing of personal data concerning

[31] OJ C […], […], p. […].
[32] OJ C […], […], p. […].

restrictions of the use of AI systems for 'real-time' remote biometric identification in publicly accessible spaces for the purpose of law enforcement, it is appropriate to base this Regulation, in as far as those specific rules are concerned, on Article 16 of the TFEU. In light of those specific rules and the recourse to Article 16 TFEU, it is appropriate to consult the European Data Protection Board.

(3) Artificial intelligence is a fast evolving family of technologies that can contribute to a wide array of economic and societal benefits across the entire spectrum of industries and social activities. By improving prediction, optimising operations and resource allocation, and personalising digital solutions available for individuals and organisations, the use of artificial intelligence can provide key competitive advantages to companies and support socially and environmentally beneficial outcomes, for example in healthcare, farming, education and training, infrastructure management, energy, transport and logistics, public services, security, justice, resource and energy efficiency, and climate change mitigation and adaptation.

(4) At the same time, depending on the circumstances regarding its specific application and use, artificial intelligence may generate risks and cause harm to public interests and rights that are protected by Union law. Such harm might be material or immaterial.

(5) A Union legal framework laying down harmonised rules on artificial intelligence is therefore needed to foster the development, use and uptake of artificial intelligence in the internal market that at the same time meets a high level of protection of public interests, such as health and safety and the protection of fundamental rights, as recognised and protected by Union law. To achieve that objective, rules regulating the placing on the market and putting into service of certain AI systems should be laid down, thus ensuring the smooth functioning of the internal market and allowing those systems to benefit from the principle of free movement of goods and services. By laying down those rules, this Regulation supports the objective of the Union of being a global leader in the development of secure, trustworthy and ethical artificial intelligence, as stated by the European Council[33], and it ensures the protection of ethical principles, as specifically requested by the European Parliament[34].

(6) The notion of AI system should be clearly defined to ensure legal certainty, while providing the flexibility to accommodate future technological developments. The definition should be based on the key functional characteristics of the software, in particular the ability, for a given set of human-defined objectives, to generate outputs such as content, predictions, recommendations, or decisions which influence the environment with which the system interacts, be it in a physical or digital dimension. AI systems can be designed to operate with varying levels of autonomy and be used on a stand-alone basis or as a component of a product, irrespective of whether the system is physically integrated into the product (embedded) or serve the functionality of the product without being integrated therein (non-embedded). The definition of AI system should be complemented by a list of specific techniques and approaches used for its development, which should be kept up-to–date in the light of market and technological

[33] European Council, Special meeting of the European Council (1 and 2 October 2020) – Conclusions, EUCO 13/20, 2020, p. 6.

[34] European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies, 2020/2012(INL).

developments through the adoption of delegated acts by the Commission to amend that list.

(7) The notion of biometric data used in this Regulation is in line with and should be interpreted consistently with the notion of biometric data as defined in Article 4(14) of Regulation (EU) 2016/679 of the European Parliament and of the Council[35], Article 3(18) of Regulation (EU) 2018/1725 of the European Parliament and of the Council[36] and Article 3(13) of Directive (EU) 2016/680 of the European Parliament and of the Council[37].

(8) The notion of remote biometric identification system as used in this Regulation should be defined functionally, as an AI system intended for the identification of natural persons at a distance through the comparison of a person's biometric data with the biometric data contained in a reference database, and without prior knowledge whether the targeted person will be present and can be identified, irrespectively of the particular technology, processes or types of biometric data used. Considering their different characteristics and manners in which they are used, as well as the different risks involved, a distinction should be made between 'real-time' and 'post' remote biometric identification systems. In the case of 'real-time' systems, the capturing of the biometric data, the comparison and the identification occur all instantaneously, near-instantaneously or in any event without a significant delay. In this regard, there should be no scope for circumventing the rules of this Regulation on the 'real-time' use of the AI systems in question by providing for minor delays. 'Real-time' systems involve the use of 'live' or 'near-'live' material, such as video footage, generated by a camera or other device with similar functionality. In the case of 'post' systems, in contrast, the biometric data have already been captured and the comparison and identification occur only after a significant delay. This involves material, such as pictures or video footage generated by closed circuit television cameras or private devices, which has been generated before the use of the system in respect of the natural persons concerned.

(9) For the purposes of this Regulation the notion of publicly accessible space should be understood as referring to any physical place that is accessible to the public, irrespective of whether the place in question is privately or publicly owned. Therefore, the notion does not cover places that are private in nature and normally not freely accessible for third parties, including law enforcement authorities, unless those parties have been specifically invited or authorised, such as homes, private clubs, offices, warehouses and factories. Online spaces are not covered either, as they are not physical spaces. However, the mere fact that certain conditions for accessing a

---

35 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 L 119, 4.5.2016, p. 1).

36 Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC (OJ L 295, 21.11.2018, p. 39)

37 Directive (EU) 2016/680 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 by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA (Law Enforcement Directive) (OJ L 119, 4.5.2016, p. 89).

particular space may apply, such as admission tickets or age restrictions, does not mean that the space is not publicly accessible within the meaning of this Regulation. Consequently, in addition to public spaces such as streets, relevant parts of government buildings and most transport infrastructure, spaces such as cinemas, theatres, shops and shopping centres are normally also publicly accessible. Whether a given space is accessible to the public should however be determined on a case-by-case basis, having regard to the specificities of the individual situation at hand.

(10) In order to ensure a level playing field and an effective protection of rights and freedoms of individuals across the Union, the rules established by this Regulation should apply to providers of AI systems in a non-discriminatory manner, irrespective of whether they are established within the Union or in a third country, and to users of AI systems established within the Union.

(11) In light of their digital nature, certain AI systems should fall within the scope of this Regulation even when they are neither placed on the market, nor put into service, nor used in the Union. This is the case for example of an operator established in the Union that contracts certain services to an operator established outside the Union in relation to an activity to be performed by an AI system that would qualify as high-risk and whose effects impact natural persons located in the Union. In those circumstances, the AI system used by the operator outside the Union could process data lawfully collected in and transferred from the Union, and provide to the contracting operator in the Union the output of that AI system resulting from that processing, without that AI system being placed on the market, put into service or used in the Union. To prevent the circumvention of this Regulation and to ensure an effective protection of natural persons located in the Union, this Regulation should also apply to providers and users of AI systems that are established in a third country, to the extent the output produced by those systems is used in the Union. Nonetheless, to take into account existing arrangements and special needs for cooperation with foreign partners with whom information and evidence is exchanged, this Regulation should not apply to public authorities of a third country and international organisations when acting in the framework of international agreements concluded at national or European level for law enforcement and judicial cooperation with the Union or with its Member States. Such agreements have been concluded bilaterally between Member States and third countries or between the European Union, Europol and other EU agencies and third countries and international organisations.

(12) This Regulation should also apply to Union institutions, offices, bodies and agencies when acting as a provider or user of an AI system. AI systems exclusively developed or used for military purposes should be excluded from the scope of this Regulation where that use falls under the exclusive remit of the Common Foreign and Security Policy regulated under Title V of the Treaty on the European Union (TEU). This Regulation should be without prejudice to the provisions regarding the liability of intermediary service providers set out in Directive 2000/31/EC of the European Parliament and of the Council [as amended by the Digital Services Act].

(13) In order to ensure a consistent and high level of protection of public interests as regards health, safety and fundamental rights, common normative standards for all high-risk AI systems should be established. Those standards should be consistent with the Charter of fundamental rights of the European Union (the Charter) and should be non-discriminatory and in line with the Union's international trade commitments.

(14) In order to introduce a proportionate and effective set of binding rules for AI systems, a clearly defined risk-based approach should be followed. That approach should tailor the type and content of such rules to the intensity and scope of the risks that AI systems can generate. It is therefore necessary to prohibit certain artificial intelligence practices, to lay down requirements for high-risk AI systems and obligations for the relevant operators, and to lay down transparency obligations for certain AI systems.

(15) Aside from the many beneficial uses of artificial intelligence, that technology can also be misused and provide novel and powerful tools for manipulative, exploitative and social control practices. Such practices are particularly harmful and should be prohibited because they contradict Union values of respect for human dignity, freedom, equality, democracy and the rule of law and Union fundamental rights, including the right to non-discrimination, data protection and privacy and the rights of the child.

(16) The placing on the market, putting into service or use of certain AI systems intended to distort human behaviour, whereby physical or psychological harms are likely to occur, should be forbidden. Such AI systems deploy subliminal components individuals cannot perceive or exploit vulnerabilities of children and people due to their age, physical or mental incapacities. They do so with the intention to materially distort the behaviour of a person and in a manner that causes or is likely to cause harm to that or another person. The intention may not be presumed if the distortion of human behaviour results from factors external to the AI system which are outside of the control of the provider or the user. Research for legitimate purposes in relation to such AI systems should not be stifled by the prohibition, if such research does not amount to use of the AI system in human-machine relations that exposes natural persons to harm and such research is carried out in accordance with recognised ethical standards for scientific research.

(17) AI systems providing social scoring of natural persons for general purpose by public authorities or on their behalf may lead to discriminatory outcomes and the exclusion of certain groups. They may violate the right to dignity and non-discrimination and the values of equality and justice. Such AI systems evaluate or classify the trustworthiness of natural persons based on their social behaviour in multiple contexts or known or predicted personal or personality characteristics. The social score obtained from such AI systems may lead to the detrimental or unfavourable treatment of natural persons or whole groups thereof in social contexts, which are unrelated to the context in which the data was originally generated or collected or to a detrimental treatment that is disproportionate or unjustified to the gravity of their social behaviour. Such AI systems should be therefore prohibited.

(18) The use of AI systems for ‘real-time’ remote biometric identification of natural persons in publicly accessible spaces for the purpose of law enforcement is considered particularly intrusive in the rights and freedoms of the concerned persons, to the extent that it may affect the private life of a large part of the population, evoke a feeling of constant surveillance and indirectly dissuade the exercise of the freedom of assembly and other fundamental rights. In addition, the immediacy of the impact and the limited opportunities for further checks or corrections in relation to the use of such systems operating in ‘real-time’ carry heightened risks for the rights and freedoms of the persons that are concerned by law enforcement activities.

(19) The use of those systems for the purpose of law enforcement should therefore be prohibited, except in three exhaustively listed and narrowly defined situations, where

the use is strictly necessary to achieve a substantial public interest, the importance of which outweighs the risks. Those situations involve the search for potential victims of crime, including missing children; certain threats to the life or physical safety of natural persons or of a terrorist attack; and the detection, localisation, identification or prosecution of perpetrators or suspects of the criminal offences referred to in Council Framework Decision 2002/584/JHA[38] if those criminal offences are punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years and as they are defined in the law of that Member State. Such threshold for the custodial sentence or detention order in accordance with national law contributes to ensure that the offence should be serious enough to potentially justify the use of 'real-time' remote biometric identification systems. Moreover, of the 32 criminal offences listed in the Council Framework Decision 2002/584/JHA, some are in practice likely to be more relevant than others, in that the recourse to 'real-time' remote biometric identification will foreseeably be necessary and proportionate to highly varying degrees for the practical pursuit of the detection, localisation, identification or prosecution of a perpetrator or suspect of the different criminal offences listed and having regard to the likely differences in the seriousness, probability and scale of the harm or possible negative consequences.

(20) In order to ensure that those systems are used in a responsible and proportionate manner, it is also important to establish that, in each of those three exhaustively listed and narrowly defined situations, certain elements should be taken into account, in particular as regards the nature of the situation giving rise to the request and the consequences of the use for the rights and freedoms of all persons concerned and the safeguards and conditions provided for with the use. In addition, the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement should be subject to appropriate limits in time and space, having regard in particular to the evidence or indications regarding the threats, the victims or perpetrator. The reference database of persons should be appropriate for each use case in each of the three situations mentioned above.

(21) Each use of a 'real-time' remote biometric identification system in publicly accessible spaces for the purpose of law enforcement should be subject to an express and specific authorisation by a judicial authority or by an independent administrative authority of a Member State. Such authorisation should in principle be obtained prior to the use, except in duly justified situations of urgency, that is, situations where the need to use the systems in question is such as to make it effectively and objectively impossible to obtain an authorisation before commencing the use. In such situations of urgency, the use should be restricted to the absolute minimum necessary and be subject to appropriate safeguards and conditions, as determined in national law and specified in the context of each individual urgent use case by the law enforcement authority itself. In addition, the law enforcement authority should in such situations seek to obtain an authorisation as soon as possible, whilst providing the reasons for not having been able to request it earlier.

(22) Furthermore, it is appropriate to provide, within the exhaustive framework set by this Regulation that such use in the territory of a Member State in accordance with this Regulation should only be possible where and in as far as the Member State in question has decided to expressly provide for the possibility to authorise such use in its

[38] Council Framework Decision 2002/584/JHA of 13 June 2002 on the European arrest warrant and the surrender procedures between Member States (OJ L 190, 18.7.2002, p. 1).

detailed rules of national law. Consequently, Member States remain free under this Regulation not to provide for such a possibility at all or to only provide for such a possibility in respect of some of the objectives capable of justifying authorised use identified in this Regulation.

(23) The use of AI systems for 'real-time' remote biometric identification of natural persons in publicly accessible spaces for the purpose of law enforcement necessarily involves the processing of biometric data. The rules of this Regulation that prohibit, subject to certain exceptions, such use, which are based on Article 16 TFEU, should apply as *lex specialis* in respect of the rules on the processing of biometric data contained in Article 10 of Directive (EU) 2016/680, thus regulating such use and the processing of biometric data involved in an exhaustive manner. Therefore, such use and processing should only be possible in as far as it is compatible with the framework set by this Regulation, without there being scope, outside that framework, for the competent authorities, where they act for purpose of law enforcement, to use such systems and process such data in connection thereto on the grounds listed in Article 10 of Directive (EU) 2016/680. In this context, this Regulation is not intended to provide the legal basis for the processing of personal data under Article 8 of Directive 2016/680. However, the use of 'real-time' remote biometric identification systems in publicly accessible spaces for purposes other than law enforcement, including by competent authorities, should not be covered by the specific framework regarding such use for the purpose of law enforcement set by this Regulation. Such use for purposes other than law enforcement should therefore not be subject to the requirement of an authorisation under this Regulation and the applicable detailed rules of national law that may give effect to it.

(24) Any processing of biometric data and other personal data involved in the use of AI systems for biometric identification, other than in connection to the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement as regulated by this Regulation, including where those systems are used by competent authorities in publicly accessible spaces for other purposes than law enforcement, should continue to comply with all requirements resulting from Article 9(1) of Regulation (EU) 2016/679, Article 10(1) of Regulation (EU) 2018/1725 and Article 10 of Directive (EU) 2016/680, as applicable.

(25) In accordance with Article 6a of Protocol No 21 on the position of the United Kingdom and Ireland in respect of the area of freedom, security and justice, as annexed to the TEU and to the TFEU, Ireland is not bound by the rules laid down in Article 5(1), point (d), (2) and (3) of this Regulation adopted on the basis of Article 16 of the TFEU which relate to the processing of personal data by the Member States when carrying out activities falling within the scope of Chapter 4 or Chapter 5 of Title V of Part Three of the TFEU, where Ireland is not bound by the rules governing the forms of judicial cooperation in criminal matters or police cooperation which require compliance with the provisions laid down on the basis of Article 16 of the TFEU.

(26) In accordance with Articles 2 and 2a of Protocol No 22 on the position of Denmark, annexed to the TEU and TFEU, Denmark is not bound by rules laid down in Article 5(1), point (d), (2) and (3) of this Regulation adopted on the basis of Article 16 of the TFEU, or subject to their application, which relate to the processing of personal data by the Member States when carrying out activities falling within the scope of Chapter 4 or Chapter 5 of Title V of Part Three of the TFEU.

(27) High-risk AI systems should only be placed on the Union market or put into service if they comply with certain mandatory requirements. Those requirements should ensure that high-risk AI systems available in the Union or whose output is otherwise used in the Union do not pose unacceptable risks to important Union public interests as recognised and protected by Union law. AI systems identified as high-risk should be limited to those that have a significant harmful impact on the health, safety and fundamental rights of persons in the Union and such limitation minimises any potential restriction to international trade, if any.

(28) AI systems could produce adverse outcomes to health and safety of persons, in particular when such systems operate as components of products. Consistently with the objectives of Union harmonisation legislation to facilitate the free movement of products in the internal market and to ensure that only safe and otherwise compliant products find their way into the market, it is important that the safety risks that may be generated by a product as a whole due to its digital components, including AI systems, are duly prevented and mitigated. For instance, increasingly autonomous robots, whether in the context of manufacturing or personal assistance and care should be able to safely operate and performs their functions in complex environments. Similarly, in the health sector where the stakes for life and health are particularly high, increasingly sophisticated diagnostics systems and systems supporting human decisions should be reliable and accurate. The extent of the adverse impact caused by the AI system on the fundamental rights protected by the Charter is of particular relevance when classifying an AI system as high-risk. Those rights include the right to human dignity, respect for private and family life, protection of personal data, freedom of expression and information, freedom of assembly and of association, and non-discrimination, consumer protection, workers' rights, rights of persons with disabilities, right to an effective remedy and to a fair trial, right of defence and the presumption of innocence, right to good administration. In addition to those rights, it is important to highlight that children have specific rights as enshrined in Article 24 of the EU Charter and in the United Nations Convention on the Rights of the Child (further elaborated in the UNCRC General Comment No. 25 as regards the digital environment), both of which require consideration of the children's vulnerabilities and provision of such protection and care as necessary for their well-being. The fundamental right to a high level of environmental protection enshrined in the Charter and implemented in Union policies should also be considered when assessing the severity of the harm that an AI system can cause, including in relation to the health and safety of persons.

(29) As regards high-risk AI systems that are safety components of products or systems, or which are themselves products or systems falling within the scope of Regulation (EC) No 300/2008 of the European Parliament and of the Council[39], Regulation (EU) No 167/2013 of the European Parliament and of the Council[40], Regulation (EU) No 168/2013 of the European Parliament and of the Council[41], Directive 2014/90/EU of

[39] Regulation (EC) No 300/2008 of the European Parliament and of the Council of 11 March 2008 on common rules in the field of civil aviation security and repealing Regulation (EC) No 2320/2002 (OJ L 97, 9.4.2008, p. 72).

[40] Regulation (EU) No 167/2013 of the European Parliament and of the Council of 5 February 2013 on the approval and market surveillance of agricultural and forestry vehicles (OJ L 60, 2.3.2013, p. 1).

[41] Regulation (EU) No 168/2013 of the European Parliament and of the Council of 15 January 2013 on the approval and market surveillance of two- or three-wheel vehicles and quadricycles (OJ L 60, 2.3.2013, p. 52).

the European Parliament and of the Council[42], Directive (EU) 2016/797 of the European Parliament and of the Council[43], Regulation (EU) 2018/858 of the European Parliament and of the Council[44], Regulation (EU) 2018/1139 of the European Parliament and of the Council[45], and Regulation (EU) 2019/2144 of the European Parliament and of the Council[46], it is appropriate to amend those acts to ensure that the Commission takes into account, on the basis of the technical and regulatory specificities of each sector, and without interfering with existing governance, conformity assessment and enforcement mechanisms and authorities established therein, the mandatory requirements for high-risk AI systems laid down in this Regulation when adopting any relevant future delegated or implementing acts on the basis of those acts.

(30) As regards AI systems that are safety components of products, or which are themselves products, falling within the scope of certain Union harmonisation legislation, it is appropriate to classify them as high-risk under this Regulation if the product in question undergoes the conformity assessment procedure with a third-party conformity assessment body pursuant to that relevant Union harmonisation legislation. In particular, such products are machinery, toys, lifts, equipment and protective systems intended for use in potentially explosive atmospheres, radio equipment, pressure equipment, recreational craft equipment, cableway installations, appliances burning gaseous fuels, medical devices, and in vitro diagnostic medical devices.

(31) The classification of an AI system as high-risk pursuant to this Regulation should not necessarily mean that the product whose safety component is the AI system, or the AI system itself as a product, is considered 'high-risk' under the criteria established in the relevant Union harmonisation legislation that applies to the product. This is notably the case for Regulation (EU) 2017/745 of the European Parliament and of the

---

42 Directive 2014/90/EU of the European Parliament and of the Council of 23 July 2014 on marine equipment and repealing Council Directive 96/98/EC (OJ L 257, 28.8.2014, p. 146).

43 Directive (EU) 2016/797 of the European Parliament and of the Council of 11 May 2016 on the interoperability of the rail system within the European Union (OJ L 138, 26.5.2016, p. 44).

44 Regulation (EU) 2018/858 of the European Parliament and of the Council of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, amending Regulations (EC) No 715/2007 and (EC) No 595/2009 and repealing Directive 2007/46/EC (OJ L 151, 14.6.2018, p. 1).

45 Regulation (EU) 2018/1139 of the European Parliament and of the Council of 4 July 2018 on common rules in the field of civil aviation and establishing a European Union Aviation Safety Agency, and amending Regulations (EC) No 2111/2005, (EC) No 1008/2008, (EU) No 996/2010, (EU) No 376/2014 and Directives 2014/30/EU and 2014/53/EU of the European Parliament and of the Council, and repealing Regulations (EC) No 552/2004 and (EC) No 216/2008 of the European Parliament and of the Council and Council Regulation (EEC) No 3922/91 (OJ L 212, 22.8.2018, p. 1).

46 Regulation (EU) 2019/2144 of the European Parliament and of the Council of 27 November 2019 on type-approval requirements for motor vehicles and their trailers, and systems, components and separate technical units intended for such vehicles, as regards their general safety and the protection of vehicle occupants and vulnerable road users, amending Regulation (EU) 2018/858 of the European Parliament and of the Council and repealing Regulations (EC) No 78/2009, (EC) No 79/2009 and (EC) No 661/2009 of the European Parliament and of the Council and Commission Regulations (EC) No 631/2009, (EU) No 406/2010, (EU) No 672/2010, (EU) No 1003/2010, (EU) No 1005/2010, (EU) No 1008/2010, (EU) No 1009/2010, (EU) No 19/2011, (EU) No 109/2011, (EU) No 458/2011, (EU) No 65/2012, (EU) No 130/2012, (EU) No 347/2012, (EU) No 351/2012, (EU) No 1230/2012 and (EU) 2015/166 (OJ L 325, 16.12.2019, p. 1).

Council[47] and Regulation (EU) 2017/746 of the European Parliament and of the Council[48], where a third-party conformity assessment is provided for medium-risk and high-risk products.

(32) As regards stand-alone AI systems, meaning high-risk AI systems other than those that are safety components of products, or which are themselves products, it is appropriate to classify them as high-risk if, in the light of their intended purpose, they pose a high risk of harm to the health and safety or the fundamental rights of persons, taking into account both the severity of the possible harm and its probability of occurrence and they are used in a number of specifically pre-defined areas specified in the Regulation. The identification of those systems is based on the same methodology and criteria envisaged also for any future amendments of the list of high-risk AI systems.

(33) Technical inaccuracies of AI systems intended for the remote biometric identification of natural persons can lead to biased results and entail discriminatory effects. This is particularly relevant when it comes to age, ethnicity, sex or disabilities. Therefore, 'real-time' and 'post' remote biometric identification systems should be classified as high-risk. In view of the risks that they pose, both types of remote biometric identification systems should be subject to specific requirements on logging capabilities and human oversight.

(34) As regards the management and operation of critical infrastructure, it is appropriate to classify as high-risk the AI systems intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity, since their failure or malfunctioning may put at risk the life and health of persons at large scale and lead to appreciable disruptions in the ordinary conduct of social and economic activities.

(35) AI systems used in education or vocational training, notably for determining access or assigning persons to educational and vocational training institutions or to evaluate persons on tests as part of or as a precondition for their education should be considered high-risk, since they may determine the educational and professional course of a person's life and therefore affect their ability to secure their livelihood. When improperly designed and used, such systems may violate the right to education and training as well as the right not to be discriminated against and perpetuate historical patterns of discrimination.

(36) AI systems used in employment, workers management and access to self-employment, notably for the recruitment and selection of persons, for making decisions on promotion and termination and for task allocation, monitoring or evaluation of persons in work-related contractual relationships, should also be classified as high-risk, since those systems may appreciably impact future career prospects and livelihoods of these persons. Relevant work-related contractual relationships should involve employees and persons providing services through platforms as referred to in the Commission Work Programme 2021. Such persons should in principle not be considered users within the meaning of this Regulation. Throughout the recruitment process and in the

[47] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (OJ L 117, 5.5.2017, p. 1).

[48] Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU (OJ L 117, 5.5.2017, p. 176).

evaluation, promotion, or retention of persons in work-related contractual relationships, such systems may perpetuate historical patterns of discrimination, for example against women, certain age groups, persons with disabilities, or persons of certain racial or ethnic origins or sexual orientation. AI systems used to monitor the performance and behaviour of these persons may also impact their rights to data protection and privacy.

(37) Another area in which the use of AI systems deserves special consideration is the access to and enjoyment of certain essential private and public services and benefits necessary for people to fully participate in society or to improve one's standard of living. In particular, AI systems used to evaluate the credit score or creditworthiness of natural persons should be classified as high-risk AI systems, since they determine those persons' access to financial resources or essential services such as housing, electricity, and telecommunication services. AI systems used for this purpose may lead to discrimination of persons or groups and perpetuate historical patterns of discrimination, for example based on racial or ethnic origins, disabilities, age, sexual orientation, or create new forms of discriminatory impacts. Considering the very limited scale of the impact and the available alternatives on the market, it is appropriate to exempt AI systems for the purpose of creditworthiness assessment and credit scoring when put into service by small-scale providers for their own use. Natural persons applying for or receiving public assistance benefits and services from public authorities are typically dependent on those benefits and services and in a vulnerable position in relation to the responsible authorities. If AI systems are used for determining whether such benefits and services should be denied, reduced, revoked or reclaimed by authorities, they may have a significant impact on persons' livelihood and may infringe their fundamental rights, such as the right to social protection, non-discrimination, human dignity or an effective remedy. Those systems should therefore be classified as high-risk. Nonetheless, this Regulation should not hamper the development and use of innovative approaches in the public administration, which would stand to benefit from a wider use of compliant and safe AI systems, provided that those systems do not entail a high risk to legal and natural persons. Finally, AI systems used to dispatch or establish priority in the dispatching of emergency first response services should also be classified as high-risk since they make decisions in very critical situations for the life and health of persons and their property.

(38) Actions by law enforcement authorities involving certain uses of AI systems are characterised by a significant degree of power imbalance and may lead to surveillance, arrest or deprivation of a natural person's liberty as well as other adverse impacts on fundamental rights guaranteed in the Charter. In particular, if the AI system is not trained with high quality data, does not meet adequate requirements in terms of its accuracy or robustness, or is not properly designed and tested before being put on the market or otherwise put into service, it may single out people in a discriminatory or otherwise incorrect or unjust manner. Furthermore, the exercise of important procedural fundamental rights, such as the right to an effective remedy and to a fair trial as well as the right of defence and the presumption of innocence, could be hampered, in particular, where such AI systems are not sufficiently transparent, explainable and documented. It is therefore appropriate to classify as high-risk a number of AI systems intended to be used in the law enforcement context where accuracy, reliability and transparency is particularly important to avoid adverse impacts, retain public trust and ensure accountability and effective redress. In view of the nature of the activities in question and the risks relating thereto, those high-risk AI systems should include in particular AI systems intended to be used by law

enforcement authorities for individual risk assessments, polygraphs and similar tools or to detect the emotional state of natural person, to detect 'deep fakes', for the evaluation of the reliability of evidence in criminal proceedings, for predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons, or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups, for profiling in the course of detection, investigation or prosecution of criminal offences, as well as for crime analytics regarding natural persons. AI systems specifically intended to be used for administrative proceedings by tax and customs authorities should not be considered high-risk AI systems used by law enforcement authorities for the purposes of prevention, detection, investigation and prosecution of criminal offences.

(39) AI systems used in migration, asylum and border control management affect people who are often in particularly vulnerable position and who are dependent on the outcome of the actions of the competent public authorities. The accuracy, non-discriminatory nature and transparency of the AI systems used in those contexts are therefore particularly important to guarantee the respect of the fundamental rights of the affected persons, notably their rights to free movement, non-discrimination, protection of private life and personal data, international protection and good administration. It is therefore appropriate to classify as high-risk AI systems intended to be used by the competent public authorities charged with tasks in the fields of migration, asylum and border control management as polygraphs and similar tools or to detect the emotional state of a natural person; for assessing certain risks posed by natural persons entering the territory of a Member State or applying for visa or asylum; for verifying the authenticity of the relevant documents of natural persons; for assisting competent public authorities for the examination of applications for asylum, visa and residence permits and associated complaints with regard to the objective to establish the eligibility of the natural persons applying for a status. AI systems in the area of migration, asylum and border control management covered by this Regulation should comply with the relevant procedural requirements set by the Directive 2013/32/EU of the European Parliament and of the Council[49], the Regulation (EC) No 810/2009 of the European Parliament and of the Council[50] and other relevant legislation.

(40) Certain AI systems intended for the administration of justice and democratic processes should be classified as high-risk, considering their potentially significant impact on democracy, rule of law, individual freedoms as well as the right to an effective remedy and to a fair trial. In particular, to address the risks of potential biases, errors and opacity, it is appropriate to qualify as high-risk AI systems intended to assist judicial authorities in researching and interpreting facts and the law and in applying the law to a concrete set of facts. Such qualification should not extend, however, to AI systems intended for purely ancillary administrative activities that do not affect the actual administration of justice in individual cases, such as anonymisation or pseudonymisation of judicial decisions, documents or data, communication between personnel, administrative tasks or allocation of resources.

---

[49] Directive 2013/32/EU of the European Parliament and of the Council of 26 June 2013 on common procedures for granting and withdrawing international protection (OJ L 180, 29.6.2013, p. 60).

[50] Regulation (EC) No 810/2009 of the European Parliament and of the Council of 13 July 2009 establishing a Community Code on Visas (Visa Code) (OJ L 243, 15.9.2009, p. 1).

(41) The fact that an AI system is classified as high risk under this Regulation should not be interpreted as indicating that the use of the system is necessarily lawful under other acts of Union law or under national law compatible with Union law, such as on the protection of personal data, on the use of polygraphs and similar tools or other systems to detect the emotional state of natural persons. Any such use should continue to occur solely in accordance with the applicable requirements resulting from the Charter and from the applicable acts of secondary Union law and national law. This Regulation should not be understood as providing for the legal ground for processing of personal data, including special categories of personal data, where relevant.

(42) To mitigate the risks from high-risk AI systems placed or otherwise put into service on the Union market for users and affected persons, certain mandatory requirements should apply, taking into account the intended purpose of the use of the system and according to the risk management system to be established by the provider.

(43) Requirements should apply to high-risk AI systems as regards the quality of data sets used, technical documentation and record-keeping, transparency and the provision of information to users, human oversight, and robustness, accuracy and cybersecurity. Those requirements are necessary to effectively mitigate the risks for health, safety and fundamental rights, as applicable in the light of the intended purpose of the system, and no other less trade restrictive measures are reasonably available, thus avoiding unjustified restrictions to trade.

(44) High data quality is essential for the performance of many AI systems, especially when techniques involving the training of models are used, with a view to ensure that the high-risk AI system performs as intended and safely and it does not become the source of discrimination prohibited by Union law. High quality training, validation and testing data sets require the implementation of appropriate data governance and management practices. Training, validation and testing data sets should be sufficiently relevant, representative and free of errors and complete in view of the intended purpose of the system. They should also have the appropriate statistical properties, including as regards the persons or groups of persons on which the high-risk AI system is intended to be used. In particular, training, validation and testing data sets should take into account, to the extent required in the light of their intended purpose, the features, characteristics or elements that are particular to the specific geographical, behavioural or functional setting or context within which the AI system is intended to be used. In order to protect the right of others from the discrimination that might result from the bias in AI systems, the providers shouldbe able to process also special categories of personal data, as a matter of substantial public interest, in order to ensure the bias monitoring, detection and correction in relation to high-risk AI systems.

(45) For the development of high-risk AI systems, certain actors, such as providers, notified bodies and other relevant entities, such as digital innovation hubs, testing experimentation facilities and researchers, should be able to access and use high quality datasets within their respective fields of activities which are related to this Regulation. European common data spaces established by the Commission and the facilitation of data sharing between businesses and with government in the public interest will be instrumental to provide trustful, accountable and non-discriminatory access to high quality data for the training, validation and testing of AI systems. For example, in health, the European health data space will facilitate non-discriminatory access to health data and the training of artificial intelligence algorithms on those datasets, in a privacy-preserving, secure, timely, transparent and trustworthy manner, and with an appropriate institutional governance. Relevant competent authorities,

including sectoral ones, providing or supporting the access to data may also support the provision of high-quality data for the training, validation and testing of AI systems.

(46) Having information on how high-risk AI systems have been developed and how they perform throughout their lifecycle is essential to verify compliance with the requirements under this Regulation. This requires keeping records and the availability of a technical documentation, containing information which is necessary to assess the compliance of the AI system with the relevant requirements. Such information should include the general characteristics, capabilities and limitations of the system, algorithms, data, training, testing and validation processes used as well as documentation on the relevant risk management system. The technical documentation should be kept up to date.

(47) To address the opacity that may make certain AI systems incomprehensible to or too complex for natural persons, a certain degree of transparency should be required for high-risk AI systems. Users should be able to interpret the system output and use it appropriately. High-risk AI systems should therefore be accompanied by relevant documentation and instructions of use and include concise and clear information, including in relation to possible risks to fundamental rights and discrimination, where appropriate.

(48) High-risk AI systems should be designed and developed in such a way that natural persons can oversee their functioning. For this purpose, appropriate human oversight measures should be identified by the provider of the system before its placing on the market or putting into service. In particular, where appropriate, such measures should guarantee that the system is subject to in-built operational constraints that cannot be overridden by the system itself and is responsive to the human operator, and that the natural persons to whom human oversight has been assigned have the necessary competence, training and authority to carry out that role.

(49) High-risk AI systems should perform consistently throughout their lifecycle and meet an appropriate level of accuracy, robustness and cybersecurity in accordance with the generally acknowledged state of the art. The level of accuracy and accuracy metrics should be communicated to the users.

(50) The technical robustness is a key requirement for high-risk AI systems. They should be resilient against risks connected to the limitations of the system (e.g. errors, faults, inconsistencies, unexpected situations) as well as against malicious actions that may compromise the security of the AI system and result in harmful or otherwise undesirable behaviour. Failure to protect against these risks could lead to safety impacts or negatively affect the fundamental rights, for example due to erroneous decisions or wrong or biased outputs generated by the AI system.

(51) Cybersecurity plays a crucial role in ensuring that AI systems are resilient against attempts to alter their use, behaviour, performance or compromise their security properties by malicious third parties exploiting the system's vulnerabilities. Cyberattacks against AI systems can leverage AI specific assets, such as training data sets (e.g. data poisoning) or trained models (e.g. adversarial attacks), or exploit vulnerabilities in the AI system's digital assets or the underlying ICT infrastructure. To ensure a level of cybersecurity appropriate to the risks, suitable measures should therefore be taken by the providers of high-risk AI systems, also taking into account as appropriate the underlying ICT infrastructure.

(52) As part of Union harmonisation legislation, rules applicable to the placing on the market, putting into service and use of high-risk AI systems should be laid down consistently with Regulation (EC) No 765/2008 of the European Parliament and of the Council[51] setting out the requirements for accreditation and the market surveillance of products, Decision No 768/2008/EC of the European Parliament and of the Council[52] on a common framework for the marketing of products and Regulation (EU) 2019/1020 of the European Parliament and of the Council[53] on market surveillance and compliance of products ('New Legislative Framework for the marketing of products').

(53) It is appropriate that a specific natural or legal person, defined as the provider, takes the responsibility for the placing on the market or putting into service of a high-risk AI system, regardless of whether that natural or legal person is the person who designed or developed the system.

(54) The provider should establish a sound quality management system, ensure the accomplishment of the required conformity assessment procedure, draw up the relevant documentation and establish a robust post-market monitoring system. Public authorities which put into service high-risk AI systems for their own use may adopt and implement the rules for the quality management system as part of the quality management system adopted at a national or regional level, as appropriate, taking into account the specificities of the sector and the competences and organisation of the public authority in question.

(55) Where a high-risk AI system that is a safety component of a product which is covered by a relevant New Legislative Framework sectorial legislation is not placed on the market or put into service independently from the product, the manufacturer of the final product as defined under the relevant New Legislative Framework legislation should comply with the obligations of the provider established in this Regulation and notably ensure that the AI system embedded in the final product complies with the requirements of this Regulation.

(56) To enable enforcement of this Regulation and create a level-playing field for operators, and taking into account the different forms of making available of digital products, it is important to ensure that, under all circumstances, a person established in the Union can provide authorities with all the necessary information on the compliance of an AI system. Therefore, prior to making their AI systems available in the Union, where an importer cannot be identified, providers established outside the Union shall, by written mandate, appoint an authorised representative established in the Union.

(57) In line with New Legislative Framework principles, specific obligations for relevant economic operators, such as importers and distributors, should be set to ensure legal certainty and facilitate regulatory compliance by those relevant operators.

---

51 Regulation (EC) No 765/2008 of the European Parliament and of the Council of 9 July 2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products and repealing Regulation (EEC) No 339/93 (OJ L 218, 13.8.2008, p. 30).

52 Decision No 768/2008/EC of the European Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products, and repealing Council Decision 93/465/EEC (OJ L 218, 13.8.2008, p. 82).

53 Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011 (Text with EEA relevance) (OJ L 169, 25.6.2019, p. 1–44).

(58) Given the nature of AI systems and the risks to safety and fundamental rights possibly associated with their use, including as regard the need to ensure proper monitoring of the performance of an AI system in a real-life setting, it is appropriate to set specific responsibilities for users. Users should in particular use high-risk AI systems in accordance with the instructions of use and certain other obligations should be provided for with regard to monitoring of the functioning of the AI systems and with regard to record-keeping, as appropriate.

(59) It is appropriate to envisage that the user of the AI system should be the natural or legal person, public authority, agency or other body under whose authority the AI system is operated except where the use is made in the course of a personal non-professional activity.

(60) In the light of the complexity of the artificial intelligence value chain, relevant third parties, notably the ones involved in the sale and the supply of software, software tools and components, pre-trained models and data, or providers of network services, should cooperate, as appropriate, with providers and users to enable their compliance with the obligations under this Regulation and with competent authorities established under this Regulation.

(61) Standardisation should play a key role to provide technical solutions to providers to ensure compliance with this Regulation. Compliance with harmonised standards as defined in Regulation (EU) No 1025/2012 of the European Parliament and of the Council[54] should be a means for providers to demonstrate conformity with the requirements of this Regulation. However, the Commission could adopt common technical specifications in areas where no harmonised standards exist or where they are insufficient.

(62) In order to ensure a high level of trustworthiness of high-risk AI systems, those systems should be subject to a conformity assessment prior to their placing on the market or putting into service.

(63) It is appropriate that, in order to minimise the burden on operators and avoid any possible duplication, for high-risk AI systems related to products which are covered by existing Union harmonisation legislation following the New Legislative Framework approach, the compliance of those AI systems with the requirements of this Regulation should be assessed as part of the conformity assessment already foreseen under that legislation. The applicability of the requirements of this Regulation should thus not affect the specific logic, methodology or general structure of conformity assessment under the relevant specific New Legislative Framework legislation. This approach is fully reflected in the interplay between this Regulation and the [Machinery Regulation]. While safety risks of AI systems ensuring safety functions in machinery are addressed by the requirements of this Regulation, certain specific requirements in the [Machinery Regulation] will ensure the safe integration of the AI system into the overall machinery, so as not to compromise the safety of the machinery as a whole.

[54] Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the Council (OJ L 316, 14.11.2012, p. 12).

The [Machinery Regulation] applies the same definition of AI system as this Regulation.

(64) Given the more extensive experience of professional pre-market certifiers in the field of product safety and the different nature of risks involved, it is appropriate to limit, at least in an initial phase of application of this Regulation, the scope of application of third-party conformity assessment for high-risk AI systems other than those related to products. Therefore, the conformity assessment of such systems should be carried out as a general rule by the provider under its own responsibility, with the only exception of AI systems intended to be used for the remote biometric identification of persons, for which the involvement of a notified body in the conformity assessment should be foreseen, to the extent they are not prohibited.

(65) In order to carry out third-party conformity assessment for AI systems intended to be used for the remote biometric identification of persons, notified bodies should be designated under this Regulation by the national competent authorities, provided they are compliant with a set of requirements, notably on independence, competence and absence of conflicts of interests.

(66) In line with the commonly established notion of substantial modification for products regulated by Union harmonisation legislation, it is appropriate that an AI system undergoes a new conformity assessment whenever a change occurs which may affect the compliance of the system with this Regulation or when the intended purpose of the system changes. In addition, as regards AI systems which continue to 'learn' after being placed on the market or put into service (i.e. they automatically adapt how functions are carried out), it is necessary to provide rules establishing that changes to the algorithm and its performance that have been pre-determined by the provider and assessed at the moment of the conformity assessment should not constitute a substantial modification.

(67) High-risk AI systems should bear the CE marking to indicate their conformity with this Regulation so that they can move freely within the internal market. Member States should not create unjustified obstacles to the placing on the market or putting into service of high-risk AI systems that comply with the requirements laid down in this Regulation and bear the CE marking.

(68) Under certain conditions, rapid availability of innovative technologies may be crucial for health and safety of persons and for society as a whole. It is thus appropriate that under exceptional reasons of public security or protection of life and health of natural persons and the protection of industrial and commercial property, Member States could authorise the placing on the market or putting into service of AI systems which have not undergone a conformity assessment.

(69) In order to facilitate the work of the Commission and the Member States in the artificial intelligence field as well as to increase the transparency towards the public, providers of high-risk AI systems other than those related to products falling within the scope of relevant existing Union harmonisation legislation, should be required to register their high-risk AI system in a EU database, to be established and managed by the Commission. The Commission should be the controller of that database, in accordance with Regulation (EU) 2018/1725 of the European Parliament and of the

Council[55]. In order to ensure the full functionality of the database, when deployed, the procedure for setting the database should include the elaboration of functional specifications by the Commission and an independent audit report.

(70) Certain AI systems intended to interact with natural persons or to generate content may pose specific risks of impersonation or deception irrespective of whether they qualify as high-risk or not. In certain circumstances, the use of these systems should therefore be subject to specific transparency obligations without prejudice to the requirements and obligations for high-risk AI systems. In particular, natural persons should be notified that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use. Moreover, natural persons should be notified when they are exposed to an emotion recognition system or a biometric categorisation system. Such information and notifications should be provided in accessible formats for persons with disabilities. Further, users, who use an AI system to generate or manipulate image, audio or video content that appreciably resembles existing persons, places or events and would falsely appear to a person to be authentic, should disclose that the content has been artificially created or manipulated by labelling the artificial intelligence output accordingly and disclosing its artificial origin.

(71) Artificial intelligence is a rapidly developing family of technologies that requires novel forms of regulatory oversight and a safe space for experimentation, while ensuring responsible innovation and integration of appropriate safeguards and risk mitigation measures. To ensure a legal framework that is innovation-friendly, future-proof and resilient to disruption, national competent authorities from one or more Member States should be encouraged to establish artificial intelligence regulatory sandboxes to facilitate the development and testing of innovative AI systems under strict regulatory oversight before these systems are placed on the market or otherwise put into service.

(72) The objectives of the regulatory sandboxes should be to foster AI innovation by establishing a controlled experimentation and testing environment in the development and pre-marketing phase with a view to ensuring compliance of the innovative AI systems with this Regulation and other relevant Union and Member States legislation; to enhance legal certainty for innovators and the competent authorities' oversight and understanding of the opportunities, emerging risks and the impacts of AI use, and to accelerate access to markets, including by removing barriers for small and medium enterprises (SMEs) and start-ups. To ensure uniform implementation across the Union and economies of scale, it is appropriate to establish common rules for the regulatory sandboxes' implementation and a framework for cooperation between the relevant authorities involved in the supervision of the sandboxes. This Regulation should provide the legal basis for the use of personal data collected for other purposes for developing certain AI systems in the public interest within the AI regulatory sandbox, in line with Article 6(4) of Regulation (EU) 2016/679, and Article 6 of Regulation (EU) 2018/1725, and without prejudice to Article 4(2) of Directive (EU) 2016/680. Participants in the sandbox should ensure appropriate safeguards and cooperate with the competent authorities, including by following their guidance and acting

[55] 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 L 119, 4.5.2016, p. 1).

expeditiously and in good faith to mitigate any high-risks to safety and fundamental rights that may arise during the development and experimentation in the sandbox. The conduct of the participants in the sandbox should be taken into account when competent authorities decide whether to impose an administrative fine under Article 83(2) of Regulation 2016/679 and Article 57 of Directive 2016/680.

(73) In order to promote and protect innovation, it is important that the interests of small-scale providers and users of AI systems are taken into particular account. To this objective, Member States should develop initiatives, which are targeted at those operators, including on awareness raising and information communication. Moreover, the specific interests and needs of small-scale providers shall be taken into account when Notified Bodies set conformity assessment fees. Translation costs related to mandatory documentation and communication with authorities may constitute a significant cost for providers and other operators, notably those of a smaller scale. Member States should possibly ensure that one of the languages determined and accepted by them for relevant providers' documentation and for communication with operators is one which is broadly understood by the largest possible number of cross-border users.

(74) In order to minimise the risks to implementation resulting from lack of knowledge and expertise in the market as well as to facilitate compliance of providers and notified bodies with their obligations under this Regulation, the AI-on demand platform, the European Digital Innovation Hubs and the Testing and Experimentation Facilities established by the Commission and the Member States at national or EU level should possibly contribute to the implementation of this Regulation. Within their respective mission and fields of competence, they may provide in particular technical and scientific support to providers and notified bodies.

(75) It is appropriate that the Commission facilitates, to the extent possible, access to Testing and Experimentation Facilities to bodies, groups or laboratories established or accredited pursuant to any relevant Union harmonisation legislation and which fulfil tasks in the context of conformity assessment of products or devices covered by that Union harmonisation legislation. This is notably the case for expert panels, expert laboratories and reference laboratories in the field of medical devices pursuant to Regulation (EU) 2017/745 and Regulation (EU) 2017/746.

(76) In order to facilitate a smooth, effective and harmonised implementation of this Regulation a European Artificial Intelligence Board should be established. The Board should be responsible for a number of advisory tasks, including issuing opinions, recommendations, advice or guidance on matters related to the implementation of this Regulation, including on technical specifications or existing standards regarding the requirements established in this Regulation and providing advice to and assisting the Commission on specific questions related to artificial intelligence.

(77) Member States hold a key role in the application and enforcement of this Regulation. In this respect, each Member State should designate one or more national competent authorities for the purpose of supervising the application and implementation of this Regulation. In order to increase organisation efficiency on the side of Member States and to set an official point of contact vis-à-vis the public and other counterparts at Member State and Union levels, in each Member State one national authority should be designated as national supervisory authority.

(78) In order to ensure that providers of high-risk AI systems can take into account the experience on the use of high-risk AI systems for improving their systems and the

design and development process or can take any possible corrective action in a timely manner, all providers should have a post-market monitoring system in place. This system is also key to ensure that the possible risks emerging from AI systems which continue to 'learn' after being placed on the market or put into service can be more efficiently and timely addressed. In this context, providers should also be required to have a system in place to report to the relevant authorities any serious incidents or any breaches to national and Union law protecting fundamental rights resulting from the use of their AI systems.

(79) In order to ensure an appropriate and effective enforcement of the requirements and obligations set out by this Regulation, which is Union harmonisation legislation, the system of market surveillance and compliance of products established by Regulation (EU) 2019/1020 should apply in its entirety. Where necessary for their mandate, national public authorities or bodies, which supervise the application of Union law protecting fundamental rights, including equality bodies, should also have access to any documentation created under this Regulation.

(80) Union legislation on financial services includes internal governance and risk management rules and requirements which are applicable to regulated financial institutions in the course of provision of those services, including when they make use of AI systems. In order to ensure coherent application and enforcement of the obligations under this Regulation and relevant rules and requirements of the Union financial services legislation, the authorities responsible for the supervision and enforcement of the financial services legislation, including where applicable the European Central Bank, should be designated as competent authorities for the purpose of supervising the implementation of this Regulation, including for market surveillance activities, as regards AI systems provided or used by regulated and supervised financial institutions. To further enhance the consistency between this Regulation and the rules applicable to credit institutions regulated under Directive 2013/36/EU of the European Parliament and of the Council[56], it is also appropriate to integrate the conformity assessment procedure and some of the providers' procedural obligations in relation to risk management, post marketing monitoring and documentation into the existing obligations and procedures under Directive 2013/36/EU. In order to avoid overlaps, limited derogations should also be envisaged in relation to the quality management system of providers and the monitoring obligation placed on users of high-risk AI systems to the extent that these apply to credit institutions regulated by Directive 2013/36/EU.

(81) The development of AI systems other than high-risk AI systems in accordance with the requirements of this Regulation may lead to a larger uptake of trustworthy artificial intelligence in the Union. Providers of non-high-risk AI systems should be encouraged to create codes of conduct intended to foster the voluntary application of the mandatory requirements applicable to high-risk AI systems. Providers should also be encouraged to apply on a voluntary basis additional requirements related, for example, to environmental sustainability, accessibility to persons with disability, stakeholders' participation in the design and development of AI systems, and diversity of the development teams. The Commission may develop initiatives, including of a sectorial

[56] Directive 2013/36/EU of the European Parliament and of the Council of 26 June 2013 on access to the activity of credit institutions and the prudential supervision of credit institutions and investment firms, amending Directive 2002/87/EC and repealing Directives 2006/48/EC and 2006/49/EC (OJ L 176, 27.6.2013, p. 338).

nature, to facilitate the lowering of technical barriers hindering cross-border exchange of data for AI development, including on data access infrastructure, semantic and technical interoperability of different types of data.

(82) It is important that AI systems related to products that are not high-risk in accordance with this Regulation and thus are not required to comply with the requirements set out herein are nevertheless safe when placed on the market or put into service. To contribute to this objective, the Directive 2001/95/EC of the European Parliament and of the Council[57] would apply as a safety net.

(83) In order to ensure trustful and constructive cooperation of competent authorities on Union and national level, all parties involved in the application of this Regulation should respect the confidentiality of information and data obtained in carrying out their tasks.

(84) Member States should take all necessary measures to ensure that the provisions of this Regulation are implemented, including by laying down effective, proportionate and dissuasive penalties for their infringement. For certain specific infringements, Member States should take into account the margins and criteria set out in this Regulation. The European Data Protection Supervisor should have the power to impose fines on Union institutions, agencies and bodies falling within the scope of this Regulation.

(85) In order to ensure that the regulatory framework can be adapted where necessary, the power to adopt acts in accordance with Article 290 TFEU should be delegated to the Commission to amend the techniques and approaches referred to in Annex I to define AI systems, the Union harmonisation legislation listed in Annex II, the high-risk AI systems listed in Annex III, the provisions regarding technical documentation listed in Annex IV, the content of the EU declaration of conformity in Annex V, the provisions regarding the conformity assessment procedures in Annex VI and VII and the provisions establishing the high-risk AI systems to which the conformity assessment procedure based on assessment of the quality management system and assessment of the technical documentation should apply. It is of particular importance that the Commission carry out appropriate consultations during its preparatory work, including at expert level, and that those consultations be conducted in accordance with the principles laid down in the Interinstitutional Agreement of 13 April 2016 on Better Law-Making[58]. In particular, to ensure equal participation in the preparation of delegated acts, the European Parliament and the Council receive all documents at the same time as Member States' experts, and their experts systematically have access to meetings of Commission expert groups dealing with the preparation of delegated acts.

(86) In order to ensure uniform conditions for the implementation of this Regulation, implementing powers should be conferred on the Commission. Those powers should be exercised in accordance with Regulation (EU) No 182/2011 of the European Parliament and of the Council[59].

(87) Since the objective of this Regulation cannot be sufficiently achieved by the Member States and can rather, by reason of the scale or effects of the action, be better achieved

---

57 Directive 2001/95/EC of the European Parliament and of the Council of 3 December 2001 on general product safety (OJ L 11, 15.1.2002, p. 4).

58 OJ L 123, 12.5.2016, p. 1.

59 Regulation (EU) No 182/2011 of the European Parliament and of the Council of 16 February 2011 laying down the rules and general principles concerning mechanisms for control by the Member States of the Commission's exercise of implementing powers (OJ L 55, 28.2.2011, p.13).

at Union level, the Union may adopt measures in accordance with the principle of subsidiarity as set out in Article 5 TEU. In accordance with the principle of proportionality as set out in that Article, this Regulation does not go beyond what is necessary in order to achieve that objective.

(88) This Regulation should apply from … [*OP – please insert the date established in Art. 85*]. However, the infrastructure related to the governance and the conformity assessment system should be operational before that date, therefore the provisions on notified bodies and governance structure should apply from … [*OP – please insert the date – three months following the entry into force of this Regulation*]. In addition, Member States should lay down and notify to the Commission the rules on penalties, including administrative fines, and ensure that they are properly and effectively implemented by the date of application of this Regulation. Therefore the provisions on penalties should apply from [*OP – please insert the date – twelve months following the entry into force of this Regulation*].

(89) The European Data Protection Supervisor and the European Data Protection Board were consulted in accordance with Article 42(2) of Regulation (EU) 2018/1725 and delivered an opinion on […]".

HAVE ADOPTED THIS REGULATION:

TITLE I

GENERAL PROVISIONS

Article 1
Subject matter

This Regulation lays down:

(a) harmonised rules for the placing on the market, the putting into service and the use of artificial intelligence systems ('AI systems') in the Union;

(a) prohibitions of certain artificial intelligence practices;

(b) specific requirements for high-risk AI systems and obligations for operators of such systems;

(c) harmonised transparency rules for AI systems intended to interact with natural persons, emotion recognition systems and biometric categorisation systems, and AI systems used to generate or manipulate image, audio or video content;

(d) rules on market monitoring and surveillance.

Article 2
Scope

1. This Regulation applies to:

(a) providers placing on the market or putting into service AI systems in the Union, irrespective of whether those providers are established within the Union or in a third country;

(b) users of AI systems located within the Union;

(c) providers and users of AI systems that are located in a third country, where the output produced by the system is used in the Union;

2. For high-risk AI systems that are safety components of products or systems, or which are themselves products or systems, falling within the scope of the following acts, only Article 84 of this Regulation shall apply:

(a) Regulation (EC) 300/2008;

(b) Regulation (EU) No 167/2013;

(c) Regulation (EU) No 168/2013;

(d) Directive 2014/90/EU;

(e) Directive (EU) 2016/797;

(f) Regulation (EU) 2018/858;

(g) Regulation (EU) 2018/1139;

(h) Regulation (EU) 2019/2144.

3. This Regulation shall not apply to AI systems developed or used exclusively for military purposes.

4. This Regulation shall not apply to public authorities in a third country nor to international organisations falling within the scope of this Regulation pursuant to paragraph 1, where those authorities or organisations use AI systems in the framework of international agreements for law enforcement and judicial cooperation with the Union or with one or more Member States.

5. This Regulation shall not affect the application of the provisions on the liability of intermediary service providers set out in Chapter II, Section IV of Directive 2000/31/EC of the European Parliament and of the Council[60] [*as to be replaced by the corresponding provisions of the Digital Services Act*].

Article 3
Definitions

For the purpose of this Regulation, the following definitions apply:

(1) 'artificial intelligence system' (AI system) means 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;

(1) 'provider' means 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;

[60] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce') (OJ L 178, 17.7.2000, p. 1).

(3) ‘small-scale provider’ means a provider that is a micro or small enterprise within the meaning of Commission Recommendation 2003/361/EC[61];

(4) ‘user’ means any natural or legal person, public authority, agency or other body using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity;

(5) ‘authorised representative’ means any natural or legal person established in the Union who has received a written mandate from a provider of an AI system to, respectively, perform and carry out on its behalf the obligations and procedures established by this Regulation;

(6) ‘importer’ means any natural or legal person established in the Union that places on the market or puts into service an AI system that bears the name or trademark of a natural or legal person established outside the Union;

(7) ‘distributor’ means any natural or legal person in the supply chain, other than the provider or the importer, that makes an AI system available on the Union market without affecting its properties;

(8) ‘operator’ means the provider, the user, the authorised representative, the importer and the distributor;

(9) ‘placing on the market’ means the first making available of an AI system on the Union market;

(10) ‘making available on the market’ means any supply of an AI system for distribution or use on the Union market in the course of a commercial activity, whether in return for payment or free of charge;

(11) ‘putting into service’ means the supply of an AI system for first use directly to the user or for own use on the Union market for its intended purpose;

(12) ‘intended purpose’ means the use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation;

(13) ‘reasonably foreseeable misuse’ means the use of an AI system in a way that is not in accordance with its intended purpose, but which may result from reasonably foreseeable human behaviour or interaction with other systems;

(14) ‘safety component of a product or system’ means a component of a product or of a system which fulfils a safety function for that product or system or the failure or malfunctioning of which endangers the health and safety of persons or property;

(15) ‘instructions for use’ means the information provided by the provider to inform the user of in particular an AI system’s intended purpose and proper use, inclusive of the specific geographical, behavioural or functional setting within which the high-risk AI system is intended to be used;

(16) ‘recall of an AI system’ means any measure aimed at achieving the return to the provider of an AI system made available to users;

---

[61] Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises (OJ L 124, 20.5.2003, p. 36).

(17) ‘withdrawal of an AI system’ means any measure aimed at preventing the distribution, display and offer of an AI system;

(18) ‘performance of an AI system’ means the ability of an AI system to achieve its intended purpose;

(19) ‘notifying authority’ means the national authority responsible for setting up and carrying out the necessary procedures for the assessment, designation and notification of conformity assessment bodies and for their monitoring;

(20) ‘conformity assessment’ means the process of verifying whether the requirements set out in Title III, Chapter 2 of this Regulation relating to an AI system have been fulfilled;

(21) ‘conformity assessment body’ means a body that performs third-party conformity assessment activities, including testing, certification and inspection;

(22) ‘notified body’ means a conformity assessment body designated in accordance with this Regulation and other relevant Union harmonisation legislation;

(23) ‘substantial modification’ means a change to the AI system following its placing on the market or putting into service which affects the compliance of the AI system with the requirements set out in Title III, Chapter 2 of this Regulation or results in a modification to the intended purpose for which the AI system has been assessed;

(24) ‘CE marking of conformity’ (CE marking) means a marking by which a provider indicates that an AI system is in conformity with the requirements set out in Title III, Chapter 2 of this Regulation and other applicable Union legislation harmonising the conditions for the marketing of products (‘Union harmonisation legislation’) providing for its affixing;

(25) ‘post-market monitoring’ means all activities carried out by providers of AI systems to proactively collect and review experience gained from the use of AI systems they place on the market or put into service for the purpose of identifying any need to immediately apply any necessary corrective or preventive actions;

(26) ‘market surveillance authority’ means the national authority carrying out the activities and taking the measures pursuant to Regulation (EU) 2019/1020;

(27) ‘harmonised standard’ means a European standard as defined in Article 2(1)(c) of Regulation (EU) No 1025/2012;

(28) ‘common specifications’ means a document, other than a standard, containing technical solutions providing a means to, comply with certain requirements and obligations established under this Regulation;

(29) ‘training data’ means data used for training an AI system through fitting its learnable parameters, including the weights of a neural network;

(30) ‘validation data’ means data used for providing an evaluation of the trained AI system and for tuning its non-learnable parameters and its learning process, among other things, in order to prevent overfitting; whereas the validation dataset can be a separate dataset or part of the training dataset, either as a fixed or variable split;

(31) ‘testing data’ means data used for providing an independent evaluation of the trained and validated AI system in order to confirm the expected performance of that system before its placing on the market or putting into service;

(32) ‘input data’ means data provided to or directly acquired by an AI system on the basis of which the system produces an output;

(33) ‘biometric data’ means personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data;

(34) ‘emotion recognition system’ means an AI system for the purpose of identifying or inferring emotions or intentions of natural persons on the basis of their biometric data;

(35) ‘biometric categorisation system’ means an AI system for the purpose of assigning natural persons to specific categories, such as sex, age, hair colour, eye colour, tattoos, ethnic origin or sexual or political orientation, on the basis of their biometric data;

(36) ‘remote biometric identification system’ means an AI system for the purpose of identifying natural persons at a distance through the comparison of a person’s biometric data with the biometric data contained in a reference database, and without prior knowledge of the user of the AI system whether the person will be present and can be identified ;

(37) ‘‘real-time’ remote biometric identification system’ means a remote biometric identification system whereby the capturing of biometric data, the comparison and the identification all occur without a significant delay. This comprises not only instant identification, but also limited short delays in order to avoid circumvention.

(38) ‘‘post’ remote biometric identification system’ means a remote biometric identification system other than a ‘real-time’ remote biometric identification system;

(39) ‘publicly accessible space’ means any physical place accessible to the public, regardless of whether certain conditions for access may apply;

(40) ‘law enforcement authority’ means:

(a) any public authority competent for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security; or

(b) any other body or entity entrusted by Member State law to exercise public authority and public powers for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security;

(41) ‘law enforcement’ means activities carried out by law enforcement authorities for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security;

(42) ‘national supervisory authority’ means the authority to which a Member State assigns the responsibility for the implementation and application of this Regulation, for coordinating the activities entrusted to that Member State, for acting as the single contact point for the Commission, and for representing the Member State at the European Artificial Intelligence Board;

(43) ‘national competent authority’ means the national supervisory authority, the notifying authority and the market surveillance authority;

(44) ‘serious incident’ means any incident that directly or indirectly leads, might have led or might lead to any of the following:

(a) the death of a person or serious damage to a person’s health, to property or the environment,

(b) a serious and irreversible disruption of the management and operation of critical infrastructure.

Article 4
Amendments to Annex I

The Commission is empowered to adopt delegated acts in accordance with Article 73 to amend the list of techniques and approaches listed in Annex I, in order to update that list to market and technological developments on the basis of characteristics that are similar to the techniques and approaches listed therein.

TITLE II

PROHIBITED ARTIFICIAL INTELLIGENCE PRACTICES

Article 5

1. The following artificial intelligence practices shall be prohibited:

(a) the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person’s consciousness in order to materially distort a person’s behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm;

(b) the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability, in order to materially distort the behaviour of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm;

(c) the placing on the market, putting into service or use of AI systems by public authorities or on their behalf for the evaluation or classification of the trustworthiness of natural persons over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics, with the social score leading to either or both of the following:

(i) detrimental or unfavourable treatment of certain natural persons or whole groups thereof in social contexts which are unrelated to the contexts in which the data was originally generated or collected;

(ii) detrimental or unfavourable treatment of certain natural persons or whole groups thereof that is unjustified or disproportionate to their social behaviour or its gravity;

(d) the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless and in as far as such use is strictly necessary for one of the following objectives:

(i) the targeted search for specific potential victims of crime, including missing children;

(ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack;

(iii) the detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence referred to in Article 2(2) of Council Framework Decision 2002/584/JHA[62] and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, as determined by the law of that Member State.

2. The use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement for any of the objectives referred to in paragraph 1 point d) shall take into account the following elements:

(a) the nature of the situation giving rise to the possible use, in particular the seriousness, probability and scale of the harm caused in the absence of the use of the system;

(b) the consequences of the use of the system for the rights and freedoms of all persons concerned, in particular the seriousness, probability and scale of those consequences.

In addition, the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement for any of the objectives referred to in paragraph 1 point d) shall comply with necessary and proportionate safeguards and conditions in relation to the use, in particular as regards the temporal, geographic and personal limitations.

3. As regards paragraphs 1, point (d) and 2, each individual use for the purpose of law enforcement of a 'real-time' remote biometric identification system in publicly accessible spaces shall be subject to a prior authorisation granted by a judicial authority or by an independent administrative authority of the Member State in which the use is to take place, issued upon a reasoned request and in accordance with the detailed rules of national law referred to in paragraph 4. However, in a duly justified situation of urgency, the use of the system may be commenced without an authorisation and the authorisation may be requested only during or after the use.

The competent judicial or administrative authority shall only grant the authorisation where it is satisfied, based on objective evidence or clear indications presented to it, that the use of the 'real-time' remote biometric identification system at issue is necessary for and proportionate to achieving one of the objectives specified in paragraph 1, point (d), as identified in the request. In deciding on the request, the competent judicial or administrative authority shall take into account the elements referred to in paragraph 2.

4. A Member State may decide to provide for the possibility to fully or partially authorise the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement within the limits and under the

[62] Council Framework Decision 2002/584/JHA of 13 June 2002 on the European arrest warrant and the surrender procedures between Member States (OJ L 190, 18.7.2002, p. 1).

conditions listed in paragraphs 1, point (d), 2 and 3. That Member State shall lay down in its national law the necessary detailed rules for the request, issuance and exercise of, as well as supervision relating to, the authorisations referred to in paragraph 3. Those rules shall also specify in respect of which of the objectives listed in paragraph 1, point (d), including which of the criminal offences referred to in point (iii) thereof, the competent authorities may be authorised to use those systems for the purpose of law enforcement.

TITLE III

HIGH-RISK AI SYSTEMS

CHAPTER 1

CLASSIFICATION OF AI SYSTEMS AS HIGH-RISK

Article 6
Classification rules for high-risk AI systems

1. Irrespective of whether an AI system is placed on the market or put into service independently from the products referred to in points (a) and (b), that AI system shall be considered high-risk where both of the following conditions are fulfilled:

   (a) the AI system is intended to be used as a safety component of a product, or is itself a product, covered by the Union harmonisation legislation listed in Annex II;

   (b) the product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment with a view to the placing on the market or putting into service of that product pursuant to the Union harmonisation legislation listed in Annex II.

2. In addition to the high-risk AI systems referred to in paragraph 1, AI systems referred to in Annex III shall also be considered high-risk.

Article 7
Amendments to Annex III

1. The Commission is empowered to adopt delegated acts in accordance with Article 73 to update the list in Annex III by adding high-risk AI systems where both of the following conditions are fulfilled:

   (a) the AI systems are intended to be used in any of the areas listed in points 1 to 8 of Annex III;

   (b) the AI systems pose a risk of harm to the health and safety, or a risk of adverse impact on fundamental rights, that is, in respect of its severity and probability of occurrence, equivalent to or greater than the risk of harm or of adverse impact posed by the high-risk AI systems already referred to in Annex III.

2. When assessing for the purposes of paragraph 1 whether an AI system poses a risk of harm to the health and safety or a risk of adverse impact on fundamental rights that is equivalent to or greater than the risk of harm posed by the high-risk AI systems

already referred to in Annex III, the Commission shall take into account the following criteria:

(a) the intended purpose of the AI system;

(b) the extent to which an AI system has been used or is likely to be used;

(c) the extent to which the use of an AI system has already caused harm to the health and safety or adverse impact on the fundamental rights or has given rise to significant concerns in relation to the materialisation of such harm or adverse impact, as demonstrated by reports or documented allegations submitted to national competent authorities;

(d) the potential extent of such harm or such adverse impact, in particular in terms of its intensity and its ability to affect a plurality of persons;

(e) the extent to which potentially harmed or adversely impacted persons are dependent on the outcome produced with an AI system, in particular because for practical or legal reasons it is not reasonably possible to opt-out from that outcome;

(f) the extent to which potentially harmed or adversely impacted persons are in a vulnerable position in relation to the user of an AI system, in particular due to an imbalance of power, knowledge, economic or social circumstances, or age;

(g) the extent to which the outcome produced with an AI system is easily reversible, whereby outcomes having an impact on the health or safety of persons shall not be considered as easily reversible;

(h) the extent to which existing Union legislation provides for:

(i) effective measures of redress in relation to the risks posed by an AI system, with the exclusion of claims for damages;

(ii) effective measures to prevent or substantially minimise those risks.

CHAPTER 2

REQUIREMENTS FOR HIGH-RISK AI SYSTEMS

Article 8
Compliance with the requirements

1. High-risk AI systems shall comply with the requirements established in this Chapter.

2. The intended purpose of the high-risk AI system and the risk management system referred to in Article 9 shall be taken into account when ensuring compliance with those requirements.

Article 9
Risk management system

1. A risk management system shall be established, implemented, documented and maintained in relation to high-risk AI systems.

2. The risk management system shall consist of a continuous iterative process run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic updating. It shall comprise the following steps:

(a) identification and analysis of the known and foreseeable risks associated with each high-risk AI system;

(b) estimation and evaluation of the risks that may emerge when the high-risk AI system is used in accordance with its intended purpose and under conditions of reasonably foreseeable misuse;

(c) evaluation of other possibly arising risks based on the analysis of data gathered from the post-market monitoring system referred to in Article 61;

(d) adoption of suitable risk management measures in accordance with the provisions of the following paragraphs.

3. The risk management measures referred to in paragraph 2, point (d) shall give due consideration to the effects and possible interactions resulting from the combined application of the requirements set out in this Chapter 2. They shall take into account the generally acknowledged state of the art, including as reflected in relevant harmonised standards or common specifications.

4. The risk management measures referred to in paragraph 2, point (d) shall be such that any residual risk associated with each hazard as well as the overall residual risk of the high-risk AI systems is judged acceptable, provided that the high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse. Those residual risks shall be communicated to the user.

In identifying the most appropriate risk management measures, the following shall be ensured:

(a) elimination or reduction of risks as far as possible through adequate design and development;

(b) where appropriate, implementation of adequate mitigation and control measures in relation to risks that cannot be eliminated;

(c) provision of adequate information pursuant to Article 13, in particular as regards the risks referred to in paragraph 2, point (b) of this Article, and, where appropriate, training to users.

In eliminating or reducing risks related to the use of the high-risk AI system, due consideration shall be given to the technical knowledge, experience, education, training to be expected by the user and the environment in which the system is intended to be used.

5. High-risk AI systems shall be tested for the purposes of identifying the most appropriate risk management measures. Testing shall ensure that high-risk AI systems perform consistently for their intended purpose and they are in compliance with the requirements set out in this Chapter.

6. Testing procedures shall be suitable to achieve the intended purpose of the AI system and do not need to go beyond what is necessary to achieve that purpose.

7. The testing of the high-risk AI systems shall be performed, as appropriate, at any point in time throughout the development process, and, in any event, prior to the placing on the market or the putting into service. Testing shall be made against preliminarily defined metrics and probabilistic thresholds that are appropriate to the intended purpose of the high-risk AI system.

8. When implementing the risk management system described in paragraphs 1 to 7, specific consideration shall be given to whether the high-risk AI system is likely to be accessed by or have an impact on children.

9. For credit institutions regulated by Directive 2013/36/EU, the aspects described in paragraphs 1 to 8 shall be part of the risk management procedures established by those institutions pursuant to Article 74 of that Directive.

Article 10
Data and data governance

1. High-risk AI systems which make use of techniques involving the training of models with data shall be developed on the basis of training, validation and testing data sets that meet the quality criteria referred to in paragraphs 2 to 5.

2. Training, validation and testing data sets shall be subject to appropriate data governance and management practices. Those practices shall concern in particular,

   (a) the relevant design choices;

   (b) data collection;

   (c) relevant data preparation processing operations, such as annotation, labelling, cleaning, enrichment and aggregation;

   (d) the formulation of relevant assumptions, notably with respect to the information that the data are supposed to measure and represent;

   (e) a prior assessment of the availability, quantity and suitability of the data sets that are needed;

   (f) examination in view of possible biases;

   (g) the identification of any possible data gaps or shortcomings, and how those gaps and shortcomings can be addressed.

3. Training, validation and testing data sets shall be relevant, representative, free of errors and complete. They shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons on which the high-risk AI system is intended to be used. These characteristics of the data sets may be met at the level of individual data sets or a combination thereof.

4. Training, validation and testing data sets shall take into account, to the extent required by the intended purpose, the characteristics or elements that are particular to the specific geographical, behavioural or functional setting within which the high-risk AI system is intended to be used.

5. To the extent that it is strictly necessary for the purposes of ensuring bias monitoring, detection and correction in relation to the high-risk AI systems, the providers of such systems may process special categories of personal data referred to in Article 9(1) of Regulation (EU) 2016/679, Article 10 of Directive (EU) 2016/680 and Article 10(1) of Regulation (EU) 2018/1725, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons, including technical limitations on the re-use and use of state-of-the-art security and privacy-preserving measures, such as pseudonymisation, or encryption where anonymisation may significantly affect the purpose pursued.

6. Appropriate data governance and management practices shall apply for the development of high-risk AI systems other than those which make use of techniques involving the training of models in order to ensure that those high-risk AI systems comply with paragraph 2.

Article 11
Technical documentation

1. The technical documentation of a high-risk AI system shall be drawn up before that system is placed on the market or put into service and shall be kept up-to date.

   The technical documentation shall be drawn up in such a way to demonstrate that the high-risk AI system complies with the requirements set out in this Chapter and provide national competent authorities and notified bodies with all the necessary information to assess the compliance of the AI system with those requirements. It shall contain, at a minimum, the elements set out in Annex IV.

2. Where a high-risk AI system related to a product, to which the legal acts listed in Annex II, section A apply, is placed on the market or put into service one single technical documentation shall be drawn up containing all the information set out in Annex IV as well as the information required under those legal acts.

3. The Commission is empowered to adopt delegated acts in accordance with Article 73 to amend Annex IV where necessary to ensure that, in the light of technical progress, the technical documentation provides all the necessary information to assess the compliance of the system with the requirements set out in this Chapter.

Article 12
Record-keeping

1. High-risk AI systems shall be designed and developed with capabilities enabling the automatic recording of events (‘logs’) while the high-risk AI systems is operating. Those logging capabilities shall conform to recognised standards or common specifications.

2. The logging capabilities shall ensure a level of traceability of the AI system’s functioning throughout its lifecycle that is appropriate to the intended purpose of the system.

3. In particular, logging capabilities shall enable the monitoring of the operation of the high-risk AI system with respect to the occurrence of situations that may result in the AI system presenting a risk within the meaning of Article 65(1) or lead to a substantial modification, and facilitate the post-market monitoring referred to in Article 61.

4. For high-risk AI systems referred to in paragraph 1, point (a) of Annex III, the logging capabilities shall provide, at a minimum:

   (a) recording of the period of each use of the system (start date and time and end date and time of each use);

   (b) the reference database against which input data has been checked by the system;

   (c) the input data for which the search has led to a match;

(d) the identification of the natural persons involved in the verification of the results, as referred to in Article 14 (5).

Article 13
Transparency and provision of information to users

1. High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately. An appropriate type and degree of transparency shall be ensured, with a view to achieving compliance with the relevant obligations of the user and of the provider set out in Chapter 3 of this Title.

2. High-risk AI systems shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise, complete, correct and clear information that is relevant, accessible and comprehensible to users.

3. The information referred to in paragraph 2 shall specify:

(a) the identity and the contact details of the provider and, where applicable, of its authorised representative;

(b) the characteristics, capabilities and limitations of performance of the high-risk AI system, including:

(i) its intended purpose;

(ii) the level of accuracy, robustness and cybersecurity referred to in Article 15 against which the high-risk AI system has been tested and validated and which can be expected, and any known and foreseeable circumstances that may have an impact on that expected level of accuracy, robustness and cybersecurity;

(iii) any known or foreseeable circumstance, related to the use of the high-risk AI system in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, which may lead to risks to the health and safety or fundamental rights;

(iv) its performance as regards the persons or groups of persons on which the system is intended to be used;

(v) when appropriate, specifications for the input data, or any other relevant information in terms of the training, validation and testing data sets used, taking into account the intended purpose of the AI system.

(c) the changes to the high-risk AI system and its performance which have been pre-determined by the provider at the moment of the initial conformity assessment, if any;

(d) the human oversight measures referred to in Article 14, including the technical measures put in place to facilitate the interpretation of the outputs of AI systems by the users;

(e) the expected lifetime of the high-risk AI system and any necessary maintenance and care measures to ensure the proper functioning of that AI system, including as regards software updates.

Article 14
Human oversight

1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use.

2. Human oversight shall aim at preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter.

3. Human oversight shall be ensured through either one or all of the following measures:

   (a) identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service;

   (b) identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the user.

4. The measures referred to in paragraph 3 shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances:

   (a) fully understand the capacities and limitations of the high-risk AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible;

   (b) remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (‘automation bias’), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;

   (c) be able to correctly interpret the high-risk AI system’s output, taking into account in particular the characteristics of the system and the interpretation tools and methods available;

   (d) be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system;

   (e) be able to intervene on the operation of the high-risk AI system or interrupt the system through a “stop” button or a similar procedure.

5. For high-risk AI systems referred to in point 1(a) of Annex III, the measures referred to in paragraph 3 shall be such as to ensure that, in addition, no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been verified and confirmed by at least two natural persons.

Article 15
Accuracy, robustness and cybersecurity

1. High-risk AI systems shall be designed and developed in such a way that they achieve, in the light of their intended purpose, an appropriate level of accuracy,

robustness and cybersecurity, and perform consistently in those respects throughout their lifecycle.

2. The levels of accuracy and the relevant accuracy metrics of high-risk AI systems shall be declared in the accompanying instructions of use.

3. High-risk AI systems shall be resilient as regards errors, faults or inconsistencies that may occur within the system or the environment in which the system operates, in particular due to their interaction with natural persons or other systems.

   The robustness of high-risk AI systems may be achieved through technical redundancy solutions, which may include backup or fail-safe plans.

   High-risk AI systems that continue to learn after being placed on the market or put into service shall be developed in such a way to ensure that possibly biased outputs due to outputs used as an input for future operations ('feedback loops') are duly addressed with appropriate mitigation measures.

4. High-risk AI systems shall be resilient as regards attempts by unauthorised third parties to alter their use or performance by exploiting the system vulnerabilities.

   The technical solutions aimed at ensuring the cybersecurity of high-risk AI systems shall be appropriate to the relevant circumstances and the risks.

   The technical solutions to address AI specific vulnerabilities shall include, where appropriate, measures to prevent and control for attacks trying to manipulate the training dataset ('data poisoning'), inputs designed to cause the model to make a mistake ('adversarial examples'), or model flaws.

CHAPTER 3

OBLIGATIONS OF PROVIDERS AND USERS OF HIGH-RISK AI SYSTEMS AND OTHER PARTIES

Article 16
Obligations of providers of high-risk AI systems

Providers of high-risk AI systems shall:

(a) ensure that their high-risk AI systems are compliant with the requirements set out in Chapter 2 of this Title;

(b) have a quality management system in place which complies with Article 17;

(c) draw-up the technical documentation of the high-risk AI system;

(d) when under their control, keep the logs automatically generated by their high-risk AI systems;

(e) ensure that the high-risk AI system undergoes the relevant conformity assessment procedure, prior to its placing on the market or putting into service;

(f) comply with the registration obligations referred to in Article 51;

(g) take the necessary corrective actions, if the high-risk AI system is not in conformity with the requirements set out in Chapter 2 of this Title;

(h) inform the national competent authorities of the Member States in which they made the AI system available or put it into service and, where applicable, the notified body of the non-compliance and of any corrective actions taken;

(i) to affix the CE marking to their high-risk AI systems to indicate the conformity with this Regulation in accordance with Article 49;

(j) upon request of a national competent authority, demonstrate the conformity of the high-risk AI system with the requirements set out in Chapter 2 of this Title.

Article 17
Quality management system

1. Providers of high-risk AI systems shall put a quality management system in place that ensures compliance with this Regulation. That system shall be documented in a systematic and orderly manner in the form of written policies, procedures and instructions, and shall include at least the following aspects:

(a) a strategy for regulatory compliance, including compliance with conformity assessment procedures and procedures for the management of modifications to the high-risk AI system;

(b) techniques, procedures and systematic actions to be used for the design, design control and design verification of the high-risk AI system;

(c) techniques, procedures and systematic actions to be used for the development, quality control and quality assurance of the high-risk AI system;

(d) examination, test and validation procedures to be carried out before, during and after the development of the high-risk AI system, and the frequency with which they have to be carried out;

(e) technical specifications, including standards, to be applied and, where the relevant harmonised standards are not applied in full, the means to be used to ensure that the high-risk AI system complies with the requirements set out in Chapter 2 of this Title;

(f) systems and procedures for data management, including data collection, data analysis, data labelling, data storage, data filtration, data mining, data aggregation, data retention and any other operation regarding the data that is performed before and for the purposes of the placing on the market or putting into service of high-risk AI systems;

(g) the risk management system referred to in Article 9;

(h) the setting-up, implementation and maintenance of a post-market monitoring system, in accordance with Article 61;

(i) procedures related to the reporting of serious incidents and of malfunctioning in accordance with Article 62;

(j) the handling of communication with national competent authorities, competent authorities, including sectoral ones, providing or supporting the access to data, notified bodies, other operators, customers or other interested parties;

(k) systems and procedures for record keeping of all relevant documentation and information;

(l) resource management, including security of supply related measures;

(m) an accountability framework setting out the responsibilities of the management and other staff with regard to all aspects listed in this paragraph.

2. The implementation of aspects referred to in paragraph 1 shall be proportionate to the size of the provider's organisation.

3. For providers that are credit institutions regulated by Directive 2013/36/ EU, the obligation to put a quality management system in place shall be deemed to be fulfilled by complying with the rules on internal governance arrangements, processes and mechanisms pursuant to Article 74 of that Directive. In that context, any harmonised standards referred to in Article 40 of this Regulation shall be taken into account.

Article 18
Obligation to draw up technical documentation

1. Providers of high-risk AI systems shall draw up the technical documentation referred to in Article 11 in accordance with Annex IV.

2. Providers that are credit institutions regulated by Directive 2013/36/EU shall maintain the technical documentation as part of the documentation concerning internal governance, arrangements, processes and mechanisms pursuant to Article 74 of that Directive.

Article 19
Conformity assessment

1. Providers of high-risk AI systems shall ensure that their systems undergo the relevant conformity assessment procedure in accordance with Article 43, prior to their placing on the market or putting into service. Where the compliance of the AI systems with the requirements set out in Chapter 2 of this Title has been demonstrated following that conformity assessment, the providers shall draw up an EU declaration of conformity in accordance with Article 48 and affix the CE marking of conformity in accordance with Article 49.

2. For high-risk AI systems referred to in point 5(b) of Annex III that are placed on the market or put into service by providers that are credit institutions regulated by Directive 2013/36/EU, the conformity assessment shall be carried out as part of the procedure referred to in Articles 97 to101 of that Directive.

Article 20
Automatically generated logs

1. Providers of high-risk AI systems shall keep the logs automatically generated by their high-risk AI systems, to the extent such logs are under their control by virtue of a contractual arrangement with the user or otherwise by law. The logs shall be kept for a period that is appropriate in the light of the intended purpose of high-risk AI system and applicable legal obligations under Union or national law.

2. Providers that are credit institutions regulated by Directive 2013/36/EU shall maintain the logs automatically generated by their high-risk AI systems as part of the documentation under Articles 74 of that Directive.

Article 21
Corrective actions

Providers of high-risk AI systems which consider or have reason to consider that a high-risk AI system which they have placed on the market or put into service is not in conformity with this Regulation shall immediately take the necessary corrective actions to bring that system into conformity, to withdraw it or to recall it, as appropriate. They shall inform the distributors of the high-risk AI system in question and, where applicable, the authorised representative and importers accordingly.

Article 22
Duty of information

Where the high-risk AI system presents a risk within the meaning of Article 65(1) and that risk is known to the provider of the system, that provider shall immediately inform the national competent authorities of the Member States in which it made the system available and, where applicable, the notified body that issued a certificate for the high-risk AI system, in particular of the non-compliance and of any corrective actions taken.

Article 23
Cooperation with competent authorities

Providers of high-risk AI systems shall, upon request by a national competent authority, provide that authority with all the information and documentation necessary to demonstrate the conformity of the high-risk AI system with the requirements set out in Chapter 2 of this Title, in an official Union language determined by the Member State concerned. Upon a reasoned request from a national competent authority, providers shall also give that authority access to the logs automatically generated by the high-risk AI system, to the extent such logs are under their control by virtue of a contractual arrangement with the user or otherwise by law.

Article 24
Obligations of product manufacturers

Where a high-risk AI system related to products to which the legal acts listed in Annex II, section A, apply, is placed on the market or put into service together with the product manufactured in accordance with those legal acts and under the name of the product manufacturer, the manufacturer of the product shall take the responsibility of the compliance of the AI system with this Regulation and, as far as the AI system is concerned, have the same obligations imposed by the present Regulation on the provider.

Article 25
Authorised representatives

1. Prior to making their systems available on the Union market, where an importer cannot be identified, providers established outside the Union shall, by written mandate, appoint an authorised representative which is established in the Union.

2. The authorised representative shall perform the tasks specified in the mandate received from the provider. The mandate shall empower the authorised representative to carry out the following tasks:

(a) keep a copy of the EU declaration of conformity and the technical documentation at the disposal of the national competent authorities and national authorities referred to in Article 63(7);

(b) provide a national competent authority, upon a reasoned request, with all the information and documentation necessary to demonstrate the conformity of a high-risk AI system with the requirements set out in Chapter 2 of this Title, including access to the logs automatically generated by the high-risk AI system to the extent such logs are under the control of the provider by virtue of a contractual arrangement with the user or otherwise by law;

(c) cooperate with competent national authorities, upon a reasoned request, on any action the latter takes in relation to the high-risk AI system.

Article 26
Obligations of importers

1. Before placing a high-risk AI system on the market, importers of such system shall ensure that:

   (a) the appropriate conformity assessment procedure has been carried out by the provider of that AI system

   (b) the provider has drawn up the technical documentation in accordance with Annex IV;

   (c) the system bears the required conformity marking and is accompanied by the required documentation and instructions of use.

2. Where an importer considers or has reason to consider that a high-risk AI system is not in conformity with this Regulation, it shall not place that system on the market until that AI system has been brought into conformity. Where the high-risk AI system presents a risk within the meaning of Article 65(1), the importer shall inform the provider of the AI system and the market surveillance authorities to that effect.

3. Importers shall indicate their name, registered trade name or registered trade mark, and the address at which they can be contacted on the high-risk AI system or, where that is not possible, on its packaging or its accompanying documentation, as applicable.

4. Importers shall ensure that, while a high-risk AI system is under their responsibility, where applicable, storage or transport conditions do not jeopardise its compliance with the requirements set out in Chapter 2 of this Title.

5. Importers shall provide national competent authorities, upon a reasoned request, with all necessary information and documentation to demonstrate the conformity of a high-risk AI system with the requirements set out in Chapter 2 of this Title in a language which can be easily understood by that national competent authority, including access to the logs automatically generated by the high-risk AI system to the extent such logs are under the control of the provider by virtue of a contractual arrangement with the user or otherwise by law. They shall also cooperate with those authorities on any action national competent authority takes in relation to that system.

Article 27
Obligations of distributors

1. Before making a high-risk AI system available on the market, distributors shall verify that the high-risk AI system bears the required CE conformity marking, that it is accompanied by the required documentation and instruction of use, and that the provider and the importer of the system, as applicable, have complied with the obligations set out in this Regulation.

2. Where a distributor considers or has reason to consider that a high-risk AI system is not in conformity with the requirements set out in Chapter 2 of this Title, it shall not make the high-risk AI system available on the market until that system has been brought into conformity with those requirements. Furthermore, where the system presents a risk within the meaning of Article 65(1), the distributor shall inform the provider or the importer of the system, as applicable, to that effect.

3. Distributors shall ensure that, while a high-risk AI system is under their responsibility, where applicable, storage or transport conditions do not jeopardise the compliance of the system with the requirements set out in Chapter 2 of this Title.

4. A distributor that considers or has reason to consider that a high-risk AI system which it has made available on the market is not in conformity with the requirements set out in Chapter 2 of this Title shall take the corrective actions necessary to bring that system into conformity with those requirements, to withdraw it or recall it or shall ensure that the provider, the importer or any relevant operator, as appropriate, takes those corrective actions. Where the high-risk AI system presents a risk within the meaning of Article 65(1), the distributor shall immediately inform the national competent authorities of the Member States in which it has made the product available to that effect, giving details, in particular, of the non-compliance and of any corrective actions taken.

5. Upon a reasoned request from a national competent authority, distributors of high-risk AI systems shall provide that authority with all the information and documentation necessary to demonstrate the conformity of a high-risk system with the requirements set out in Chapter 2 of this Title. Distributors shall also cooperate with that national competent authority on any action taken by that authority.

Article 28
Obligations of distributors, importers, users or any other third-party

1. Any distributor, importer, user or other third-party shall be considered a provider for the purposes of this Regulation and shall be subject to the obligations of the provider under Article 16, in any of the following circumstances:

   (a) they place on the market or put into service a high-risk AI system under their name or trademark;

   (b) they modify the intended purpose of a high-risk AI system already placed on the market or put into service;

   (c) they make a substantial modification to the high-risk AI system.

2. Where the circumstances referred to in paragraph 1, point (b) or (c), occur, the provider that initially placed the high-risk AI system on the market or put it into service shall no longer be considered a provider for the purposes of this Regulation.

Article 29
Obligations of users of high-risk AI systems

1. Users of high-risk AI systems shall use such systems in accordance with the instructions of use accompanying the systems, pursuant to paragraphs 2 and 5.

2. The obligations in paragraph 1 are without prejudice to other user obligations under Union or national law and to the user's discretion in organising its own resources and activities for the purpose of implementing the human oversight measures indicated by the provider.

3. Without prejudice to paragraph 1, to the extent the user exercises control over the input data, that user shall ensure that input data is relevant in view of the intended purpose of the high-risk AI system.

4. Users shall monitor the operation of the high-risk AI system on the basis of the instructions of use. When they have reasons to consider that the use in accordance with the instructions of use may result in the AI system presenting a risk within the meaning of Article 65(1) they shall inform the provider or distributor and suspend the use of the system. They shall also inform the provider or distributor when they have identified any serious incident or any malfunctioning within the meaning of Article 62 and interrupt the use of the AI system. In case the user is not able to reach the provider, Article 62 shall apply mutatis mutandis.

   For users that are credit institutions regulated by Directive 2013/36/EU, the monitoring obligation set out in the first subparagraph shall be deemed to be fulfilled by complying with the rules on internal governance arrangements, processes and mechanisms pursuant to Article 74 of that Directive.

5. Users of high-risk AI systems shall keep the logs automatically generated by that high-risk AI system, to the extent such logs are under their control. The logs shall be kept for a period that is appropriate in the light of the intended purpose of the high-risk AI system and applicable legal obligations under Union or national law.

   Users that are credit institutions regulated by Directive 2013/36/EU shall maintain the logs as part of the documentation concerning internal governance arrangements, processes and mechanisms pursuant to Article 74 of that Directive.

6. Users of high-risk AI systems shall use the information provided under Article 13 to comply with their obligation to carry out a data protection impact assessment under Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680, where applicable.

CHAPTER 4

NOTIFIYING AUTHORITIES AND NOTIFIED BODIES

Article 30
Notifying authorities

1. Each Member State shall designate or establish a notifying authority responsible for setting up and carrying out the necessary procedures for the assessment, designation and notification of conformity assessment bodies and for their monitoring.

2. Member States may designate a national accreditation body referred to in Regulation (EC) No 765/2008 as a notifying authority.

3. Notifying authorities shall be established, organised and operated in such a way that no conflict of interest arises with conformity assessment bodies and the objectivity and impartiality of their activities are safeguarded.

4. Notifying authorities shall be organised in such a way that decisions relating to the notification of conformity assessment bodies are taken by competent persons different from those who carried out the assessment of those bodies.

5. Notifying authorities shall not offer or provide any activities that conformity assessment bodies perform or any consultancy services on a commercial or competitive basis.

6. Notifying authorities shall safeguard the confidentiality of the information they obtain.

7. Notifying authorities shall have a sufficient number of competent personnel at their disposal for the proper performance of their tasks.

8. Notifying authorities shall make sure that conformity assessments are carried out in a proportionate manner, avoiding unnecessary burdens for providers and that notified bodies perform their activities taking due account of the size of an undertaking, the sector in which it operates, its structure and the degree of complexity of the AI system in question.

Article 31
Application of a conformity assessment body for notification

1. Conformity assessment bodies shall submit an application for notification to the notifying authority of the Member State in which they are established.

2. The application for notification shall be accompanied by a description of the conformity assessment activities, the conformity assessment module or modules and the artificial intelligence technologies for which the conformity assessment body claims to be competent, as well as by an accreditation certificate, where one exists, issued by a national accreditation body attesting that the conformity assessment body fulfils the requirements laid down in Article 33. Any valid document related to existing designations of the applicant notified body under any other Union harmonisation legislation shall be added.

3. Where the conformity assessment body concerned cannot provide an accreditation certificate, it shall provide the notifying authority with the documentary evidence necessary for the verification, recognition and regular monitoring of its compliance with the requirements laid down in Article 33. For notified bodies which are designated under any other Union harmonisation legislation, all documents and certificates linked to those designations may be used to support their designation procedure under this Regulation, as appropriate.

Article 32
Notification procedure

1. Notifying authorities may notify only conformity assessment bodies which have satisfied the requirements laid down in Article 33.

2. Notifying authorities shall notify the Commission and the other Member States using the electronic notification tool developed and managed by the Commission.

3. The notification shall include full details of the conformity assessment activities, the conformity assessment module or modules and the artificial intelligence technologies concerned.

4. The conformity assessment body concerned may perform the activities of a notified body only where no objections are raised by the Commission or the other Member States within one month of a notification.

5. Notifying authorities shall notify the Commission and the other Member States of any subsequent relevant changes to the notification.

Article 33
Notified bodies

1. Notified bodies shall verify the conformity of high-risk AI system in accordance with the conformity assessment procedures referred to in Article 43.

2. Notified bodies shall satisfy the organisational, quality management, resources and process requirements that are necessary to fulfil their tasks.

3. The organisational structure, allocation of responsibilities, reporting lines and operation of notified bodies shall be such as to ensure that there is confidence in the performance by and in the results of the conformity assessment activities that the notified bodies conduct.

4. Notified bodies shall be independent of the provider of a high-risk AI system in relation to which it performs conformity assessment activities. Notified bodies shall also be independent of any other operator having an economic interest in the high-risk AI system that is assessed, as well as of any competitors of the provider.

5. Notified bodies shall be organised and operated so as to safeguard the independence, objectivity and impartiality of their activities. Notified bodies shall document and implement a structure and procedures to safeguard impartiality and to promote and apply the principles of impartiality throughout their organisation, personnel and assessment activities.

6. Notified bodies shall have documented procedures in place ensuring that their personnel, committees, subsidiaries, subcontractors and any associated body or personnel of external bodies respect the confidentiality of the information which comes into their possession during the performance of conformity assessment activities, except when disclosure is required by law. The staff of notified bodies shall be bound to observe professional secrecy with regard to all information obtained in carrying out their tasks under this Regulation, except in relation to the notifying authorities of the Member State in which their activities are carried out.

7. Notified bodies shall have procedures for the performance of activities which take due account of the size of an undertaking, the sector in which it operates, its structure, the degree of complexity of the AI system in question.

8. Notified bodies shall take out appropriate liability insurance for their conformity assessment activities, unless liability is assumed by the Member State concerned in accordance with national law or that Member State is directly responsible for the conformity assessment.

9. Notified bodies shall be capable of carrying out all the tasks falling to them under this Regulation with the highest degree of professional integrity and the requisite

competence in the specific field, whether those tasks are carried out by notified bodies themselves or on their behalf and under their responsibility.

10. Notified bodies shall have sufficient internal competences to be able to effectively evaluate the tasks conducted by external parties on their behalf. To that end, at all times and for each conformity assessment procedure and each type of high-risk AI system in relation to which they have been designated, the notified body shall have permanent availability of sufficient administrative, technical and scientific personnel who possess experience and knowledge relating to the relevant artificial intelligence technologies, data and data computing and to the requirements set out in Chapter 2 of this Title.

11. Notified bodies shall participate in coordination activities as referred to in Article 38. They shall also take part directly or be represented in European standardisation organisations, or ensure that they are aware and up to date in respect of relevant standards.

12. Notified bodies shall make available and submit upon request all relevant documentation, including the providers' documentation, to the notifying authority referred to in Article 30 to allow it to conduct its assessment, designation, notification, monitoring and surveillance activities and to facilitate the assessment outlined in this Chapter.

Article 34
Subsidiaries of and subcontracting by notified bodies

1. Where a notified body subcontracts specific tasks connected with the conformity assessment or has recourse to a subsidiary, it shall ensure that the subcontractor or the subsidiary meets the requirements laid down in Article 33 and shall inform the notifying authority accordingly.

2. Notified bodies shall take full responsibility for the tasks performed by subcontractors or subsidiaries wherever these are established.

3. Activities may be subcontracted or carried out by a subsidiary only with the agreement of the provider.

4. Notified bodies shall keep at the disposal of the notifying authority the relevant documents concerning the assessment of the qualifications of the subcontractor or the subsidiary and the work carried out by them under this Regulation.

Article 35
Identification numbers and lists of notified bodies designated under this Regulation

1. The Commission shall assign an identification number to notified bodies. It shall assign a single number, even where a body is notified under several Union acts.

2. The Commission shall make publicly available the list of the bodies notified under this Regulation, including the identification numbers that have been assigned to them and the activities for which they have been notified. The Commission shall ensure that the list is kept up to date.

Article 36
Changes to notifications

1. Where a notifying authority has suspicions or has been informed that a notified body no longer meets the requirements laid down in Article 33, or that it is failing to fulfil its obligations, that authority shall without delay investigate the matter with the utmost diligence. In that context, it shall inform the notified body concerned about the objections raised and give it the possibility to make its views known. If the notifying authority comes to the conclusion that the notified body investigation no longer meets the requirements laid down in Article 33 or that it is failing to fulfil its obligations, it shall restrict, suspend or withdraw the notification as appropriate, depending on the seriousness of the failure. It shall also immediately inform the Commission and the other Member States accordingly.

2. In the event of restriction, suspension or withdrawal of notification, or where the notified body has ceased its activity, the notifying authority shall take appropriate steps to ensure that the files of that notified body are either taken over by another notified body or kept available for the responsible notifying authorities at their request.

Article 37
Challenge to the competence of notified bodies

1. The Commission shall, where necessary, investigate all cases where there are reasons to doubt whether a notified body complies with the requirements laid down in Article 33.

2. The Notifying authority shall provide the Commission, on request, with all relevant information relating to the notification of the notified body concerned.

3. The Commission shall ensure that all confidential information obtained in the course of its investigations pursuant to this Article is treated confidentially.

4. Where the Commission ascertains that a notified body does not meet or no longer meets the requirements laid down in Article 33, it shall adopt a reasoned decision requesting the notifying Member State to take the necessary corrective measures, including withdrawal of notification if necessary. That implementing act shall be adopted in accordance with the examination procedure referred to in Article 74(2).

Article 38
Coordination of notified bodies

1. The Commission shall ensure that, with regard to the areas covered by this Regulation, appropriate coordination and cooperation between notified bodies active in the conformity assessment procedures of AI systems pursuant to this Regulation are put in place and properly operated in the form of a sectoral group of notified bodies**.**

2. Member States shall ensure that the bodies notified by them participate in the work of that group, directly or by means of designated representatives.

Article 39
Conformity assessment bodies of third countries

Conformity assessment bodies established under the law of a third country with which the Union has concluded an agreement may be authorised to carry out the activities of notified Bodies under this Regulation.

CHAPTER 5

STANDARDS, CONFORMITY ASSESSMENT, CERTIFICATES, REGISTRATION

Article 40
Harmonised standards

High-risk AI systems which are in conformity with harmonised standards or parts thereof the references of which have been published in the Official Journal of the European Union shall be presumed to be in conformity with the requirements set out in Chapter 2 of this Title, to the extent those standards cover those requirements.

Article 41
Common specifications

1. Where harmonised standards referred to in Article 40 do not exist or where the Commission considers that the relevant harmonised standards are insufficient or that there is a need to address specific safety or fundamental right concerns, the Commission may, by means of implementing acts, adopt common specifications in respect of the requirements set out in Chapter 2 of this Title. Those implementing acts shall be adopted in accordance with the examination procedure referred to in Article 74(2).
2. The Commission, when preparing the common specifications referred to in paragraph 1, shall gather the views of relevant bodies or expert groups established under relevant sectorial Union law.
3. High-risk AI systems which are in conformity with the common specifications referred to in paragraph 1 shall be presumed to be in conformity with the requirements set out in Chapter 2 of this Title, to the extent those common specifications cover those requirements.
4. Where providers do not comply with the common specifications referred to in paragraph 1, they shall duly justify that they have adopted technical solutions that are at least equivalent thereto.

Article 42
Presumption of conformity with certain requirements

1. Taking into account their intended purpose, high-risk AI systems that have been trained and tested on data concerning the specific geographical, behavioural and functional setting within which they are intended to be used shall be presumed to be in compliance with the requirement set out in Article 10(4).

2. High-risk AI systems that have been certified or for which a statement of conformity has been issued under a cybersecurity scheme pursuant to Regulation (EU) 2019/881 of the European Parliament and of the Council[63] and the references of which have been published in the Official Journal of the European Union shall be presumed to be in compliance with the cybersecurity requirements set out in Article 15 of this Regulation in so far as the cybersecurity certificate or statement of conformity or parts thereof cover those requirements.

Article 43
Conformity assessment

1. For high-risk AI systems listed in point 1 of Annex III, where, in demonstrating the compliance of a high-risk AI system with the requirements set out in Chapter 2 of this Title, the provider has applied harmonised standards referred to in Article 40, or, where applicable, common specifications referred to in Article 41, the provider shall follow one of the following procedures:

   (a) the conformity assessment procedure based on internal control referred to in Annex VI;

   (b) the conformity assessment procedure based on assessment of the quality management system and assessment of the technical documentation, with the involvement of a notified body, referred to in Annex VII.

   Where, in demonstrating the compliance of a high-risk AI system with the requirements set out in Chapter 2 of this Title, the provider has not applied or has applied only in part harmonised standards referred to in Article 40, or where such harmonised standards do not exist and common specifications referred to in Article 41 are not available, the provider shall follow the conformity assessment procedure set out in Annex VII.

   For the purpose of the conformity assessment procedure referred to in Annex VII, the provider may choose any of the notified bodies. However, when the system is intended to be put into service by law enforcement, immigration or asylum authorities as well as EU institutions, bodies or agencies, the market surveillance authority referred to in Article 63(5) or (6), as applicable, shall act as a notified body.

2. For high-risk AI systems referred to in points 2 to 8 of Annex III, providers shall follow the conformity assessment procedure based on internal control as referred to in Annex VI, which does not provide for the involvement of a notified body. For high-risk AI systems referred to in point 5(b) of Annex III, placed on the market or put into service by credit institutions regulated by Directive 2013/36/EU, the conformity assessment shall be carried out as part of the procedure referred to in Articles 97 to101 of that Directive.

3. For high-risk AI systems, to which legal acts listed in Annex II, section A, apply, the provider shall follow the relevant conformity assessment as required under those legal acts. The requirements set out in Chapter 2 of this Title shall apply to those

[63] Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act) (OJ L 151, 7.6.2019, p. 1).

high-risk AI systems and shall be part of that assessment. Points 4.3., 4.4., 4.5. and the fifth paragraph of point 4.6 of Annex VII shall also apply.

For the purpose of that assessment, notified bodies which have been notified under those legal acts shall be entitled to control the conformity of the high-risk AI systems with the requirements set out in Chapter 2 of this Title, provided that the compliance of those notified bodies with requirements laid down in Article 33(4), (9) and (10) has been assessed in the context of the notification procedure under those legal acts.

Where the legal acts listed in Annex II, section A, enable the manufacturer of the product to opt out from a third-party conformity assessment, provided that that manufacturer has applied all harmonised standards covering all the relevant requirements, that manufacturer may make use of that option only if he has also applied harmonised standards or, where applicable, common specifications referred to in Article 41, covering the requirements set out in Chapter 2 of this Title.

4. High-risk AI systems shall undergo a new conformity assessment procedure whenever they are substantially modified, regardless of whether the modified system is intended to be further distributed or continues to be used by the current user.

   For high-risk AI systems that continue to learn after being placed on the market or put into service, changes to the high-risk AI system and its performance that have been pre-determined by the provider at the moment of the initial conformity assessment and are part of the information contained in the technical documentation referred to in point 2(f) of Annex IV, shall not constitute a substantial modification.

5. The Commission is empowered to adopt delegated acts in accordance with Article 73 for the purpose of updating Annexes VI and Annex VII in order to introduce elements of the conformity assessment procedures that become necessary in light of technical progress.

6. The Commission is empowered to adopt delegated acts to amend paragraphs 1 and 2 in order to subject high-risk AI systems referred to in points 2 to 8 of Annex III to the conformity assessment procedure referred to in Annex VII or parts thereof. The Commission shall adopt such delegated acts taking into account the effectiveness of the conformity assessment procedure based on internal control referred to in Annex VI in preventing or minimizing the risks to health and safety and protection of fundamental rights posed by such systems as well as the availability of adequate capacities and resources among notified bodies.

Article 44
Certificates

1. Certificates issued by notified bodies in accordance with Annex VII shall be drawn-up in an official Union language determined by the Member State in which the notified body is established or in an official Union language otherwise acceptable to the notified body.

2. Certificates shall be valid for the period they indicate, which shall not exceed five years. On application by the provider, the validity of a certificate may be extended for further periods, each not exceeding five years, based on a re-assessment in accordance with the applicable conformity assessment procedures.

3. Where a notified body finds that an AI system no longer meets the requirements set out in Chapter 2 of this Title, it shall, taking account of the principle of

proportionality, suspend or withdraw the certificate issued or impose any restrictions on it, unless compliance with those requirements is ensured by appropriate corrective action taken by the provider of the system within an appropriate deadline set by the notified body. The notified body shall give reasons for its decision.

Article 45
Appeal against decisions of notified bodies

Member States shall ensure that an appeal procedure against decisions of the notified bodies is available to parties having a legitimate interest in that decision.

Article 46
Information obligations of notified bodies

1. Notified bodies shall inform the notifying authority of the following:
   (a) any Union technical documentation assessment certificates, any supplements to those certificates, quality management system approvals issued in accordance with the requirements of Annex VII;
   (b) any refusal, restriction, suspension or withdrawal of a Union technical documentation assessment certificate or a quality management system approval issued in accordance with the requirements of Annex VII;
   (c) any circumstances affecting the scope of or conditions for notification;
   (d) any request for information which they have received from market surveillance authorities regarding conformity assessment activities;
   (e) on request, conformity assessment activities performed within the scope of their notification and any other activity performed, including cross-border activities and subcontracting.
2. Each notified body shall inform the other notified bodies of:
   (a) quality management system approvals which it has refused, suspended or withdrawn, and, upon request, of quality system approvals which it has issued;
   (b) EU technical documentation assessment certificates or any supplements thereto which it has refused, withdrawn, suspended or otherwise restricted, and, upon request, of the certificates and/or supplements thereto which it has issued.
3. Each notified body shall provide the other notified bodies carrying out similar conformity assessment activities covering the same artificial intelligence technologies with relevant information on issues relating to negative and, on request, positive conformity assessment results.

Article 47
Derogation from conformity assessment procedure

1. By way of derogation from Article 43, any market surveillance authority may authorise the placing on the market or putting into service of specific high-risk AI systems within the territory of the Member State concerned, for exceptional reasons of public security or the protection of life and health of persons, environmental protection and the protection of key industrial and infrastructural assets. That authorisation shall be for a limited period of time, while the necessary conformity

assessment procedures are being carried out, and shall terminate once those procedures have been completed. The completion of those procedures shall be undertaken without undue delay.

2. The authorisation referred to in paragraph 1 shall be issued only if the market surveillance authority concludes that the high-risk AI system complies with the requirements of Chapter 2 of this Title. The market surveillance authority shall inform the Commission and the other Member States of any authorisation issued pursuant to paragraph 1.

3. Where, within 15 calendar days of receipt of the information referred to in paragraph 2, no objection has been raised by either a Member State or the Commission in respect of an authorisation issued by a market surveillance authority of a Member State in accordance with paragraph 1, that authorisation shall be deemed justified.

4. Where, within 15 calendar days of receipt of the notification referred to in paragraph 2, objections are raised by a Member State against an authorisation issued by a market surveillance authority of another Member State, or where the Commission considers the authorisation to be contrary to Union law or the conclusion of the Member States regarding the compliance of the system as referred to in paragraph 2 to be unfounded, the Commission shall without delay enter into consultation with the relevant Member State; the operator(s) concerned shall be consulted and have the possibility to present their views. In view thereof, the Commission shall decide whether the authorisation is justified or not. The Commission shall address its decision to the Member State concerned and the relevant operator or operators.

5. If the authorisation is considered unjustified, this shall be withdrawn by the market surveillance authority of the Member State concerned.

6. By way of derogation from paragraphs 1 to 5, for high-risk AI systems intended to be used as safety components of devices, or which are themselves devices, covered by Regulation (EU) 2017/745 and Regulation (EU) 2017/746, Article 59 of Regulation (EU) 2017/745 and Article 54 of Regulation (EU) 2017/746 shall apply also with regard to the derogation from the conformity assessment of the compliance with the requirements set out in Chapter 2 of this Title.

Article 48
EU declaration of conformity

1. The provider shall draw up a written EU declaration of conformity for each AI system and keep it at the disposal of the national competent authorities for 10 years after the AI system has been placed on the market or put into service. The EU declaration of conformity shall identify the AI system for which it has been drawn up. A copy of the EU declaration of conformity shall be given to the relevant national competent authorities upon request.

2. The EU declaration of conformity shall state that the high-risk AI system in question meets the requirements set out in Chapter 2 of this Title. The EU declaration of conformity shall contain the information set out in Annex V and shall be translated into an official Union language or languages required by the Member State(s) in which the high-risk AI system is made available.

3. Where high-risk AI systems are subject to other Union harmonisation legislation which also requires an EU declaration of conformity, a single EU declaration of conformity shall be drawn up in respect of all Union legislations applicable to the

high-risk AI system. The declaration shall contain all the information required for identification of the Union harmonisation legislation to which the declaration relates.

4. By drawing up the EU declaration of conformity, the provider shall assume responsibility for compliance with the requirements set out in Chapter 2 of this Title. The provider shall keep the EU declaration of conformity up-to-date as appropriate.

5. The Commission shall be empowered to adopt delegated acts in accordance with Article 73 for the purpose of updating the content of the EU declaration of conformity set out in Annex V in order to introduce elements that become necessary in light of technical progress.

Article 49
CE marking of conformity

1. The CE marking shall be affixed visibly, legibly and indelibly for high-risk AI systems. Where that is not possible or not warranted on account of the nature of the high-risk AI system, it shall be affixed to the packaging or to the accompanying documentation, as appropriate.

2. The CE marking referred to in paragraph 1 of this Article shall be subject to the general principles set out in Article 30 of Regulation (EC) No 765/2008.

3. Where applicable, the CE marking shall be followed by the identification number of the notified body responsible for the conformity assessment procedures set out in Article 43. The identification number shall also be indicated in any promotional material which mentions that the high-risk AI system fulfils the requirements for CE marking.

Article 50
Document retention

The provider shall, for a period ending 10 years after the AI system has been placed on the market or put into service, keep at the disposal of the national competent authorities:

(a) the technical documentation referred to in Article 11;

(b) the documentation concerning the quality management system referred to Article 17;

(c) the documentation concerning the changes approved by notified bodies where applicable;

(d) the decisions and other documents issued by the notified bodies where applicable;

(e) the EU declaration of conformity referred to in Article 48.

Article 51
Registration

Before placing on the market or putting into service a high-risk AI system referred to in Article 6(2), the provider or, where applicable, the authorised representative shall register that system in the EU database referred to in Article 60.

TITLE IV

TRANSPARENCY OBLIGATIONS FOR CERTAIN AI SYSTEMS

Article 52
Transparency obligations for certain AI systems

1. Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate and prosecute criminal offences, unless those systems are available for the public to report a criminal offence.

2. Users of an emotion recognition system or a biometric categorisation system shall inform of the operation of the system the natural persons exposed thereto. This obligation shall not apply to AI systems used for biometric categorisation, which are permitted by law to detect, prevent and investigate criminal offences.

3. Users of an AI system that generates or manipulates image, 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 (‘deep fake’), shall disclose that the content has been artificially generated or manipulated.

   However, the first subparagraph shall not apply where the use is authorised by law to detect, prevent, investigate and prosecute criminal offences or it is necessary for the exercise of the right to freedom of expression and the right to freedom of the arts and sciences guaranteed in the Charter of Fundamental Rights of the EU, and subject to appropriate safeguards for the rights and freedoms of third parties.

4. Paragraphs 1, 2 and 3 shall not affect the requirements and obligations set out in Title III of this Regulation.

TITLE V

MEASURES IN SUPPORT OF INNOVATION

Article 53
AI regulatory sandboxes

1. AI regulatory sandboxes established by one or more Member States competent authorities or the European Data Protection Supervisor shall provide a controlled environment that facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market or putting into service pursuant to a specific plan. This shall take place under the direct supervision and guidance by the competent authorities with a view to ensuring compliance with the requirements of this Regulation and, where relevant, other Union and Member States legislation supervised within the sandbox.

2. Member States shall ensure that to the extent the innovative AI systems involve the processing of personal data or otherwise fall under the supervisory remit of other national authorities or competent authorities providing or supporting access to data,

the national data protection authorities and those other national authorities are associated to the operation of the AI regulatory sandbox.

3. The AI regulatory sandboxes shall not affect the supervisory and corrective powers of the competent authorities. Any significant risks to health and safety and fundamental rights identified during the development and testing of such systems shall result in immediate mitigation and, failing that, in the suspension of the development and testing process until such mitigation takes place.

4. Participants in the AI regulatory sandbox shall remain liable under applicable Union and Member States liability legislation for any harm inflicted on third parties as a result from the experimentation taking place in the sandbox.

5. Member States' competent authorities that have established AI regulatory sandboxes shall coordinate their activities and cooperate within the framework of the European Artificial Intelligence Board. They shall submit annual reports to the Board and the Commission on the results from the implementation of those scheme, including good practices, lessons learnt and recommendations on their setup and, where relevant, on the application of this Regulation and other Union legislation supervised within the sandbox.

6. The modalities and the conditions of the operation of the AI regulatory sandboxes, including the eligibility criteria and the procedure for the application, selection, participation and exiting from the sandbox, and the rights and obligations of the participants shall be set out in implementing acts. Those implementing acts shall be adopted in accordance with the examination procedure referred to in Article 74(2).

Article 54

Further processing of personal data for developing certain AI systems in the public interest in the AI regulatory sandbox

1. In the AI regulatory sandbox personal data lawfully collected for other purposes shall be processed for the purposes of developing and testing certain innovative AI systems in the sandbox under the following conditions:

   (a) the innovative AI systems shall be developed for safeguarding substantial public interest in one or more of the following areas:

      (i) the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security, under the control and responsibility of the competent authorities. The processing shall be based on Member State or Union law;

      (ii) public safety and public health, including disease prevention, control and treatment;

      (iii) a high level of protection and improvement of the quality of the environment;

   (b) the data processed are necessary for complying with one or more of the requirements referred to in Title III, Chapter 2 where those requirements cannot be effectively fulfilled by processing anonymised, synthetic or other non-personal data;

(c) there are effective monitoring mechanisms to identify if any high risks to the fundamental rights of the data subjects may arise during the sandbox experimentation as well as response mechanism to promptly mitigate those risks and, where necessary, stop the processing;

(d) any personal data to be processed in the context of the sandbox are in a functionally separate, isolated and protected data processing environment under the control of the participants and only authorised persons have access to that data;

(e) any personal data processed are not be transmitted, transferred or otherwise accessed by other parties;

(f) any processing of personal data in the context of the sandbox do not lead to measures or decisions affecting the data subjects;

(g) any personal data processed in the context of the sandbox are deleted once the participation in the sandbox has terminated or the personal data has reached the end of its retention period;

(h) the logs of the processing of personal data in the context of the sandbox are kept for the duration of the participation in the sandbox and 1 year after its termination, solely for the purpose of and only as long as necessary for fulfilling accountability and documentation obligations under this Article or other application Union or Member States legislation;

(i) complete and detailed description of the process and rationale behind the training, testing and validation of the AI system is kept together with the testing results as part of the technical documentation in Annex IV;

(j) a short summary of the AI project developed in the sandbox, its objectives and expected results published on the website of the competent authorities.

2. Paragraph 1 is without prejudice to Union or Member States legislation excluding processing for other purposes than those explicitly mentioned in that legislation.

Article 55
Measures for small-scale providers and users

1. Member States shall undertake the following actions:

(a) provide small-scale providers and start-ups with priority access to the AI regulatory sandboxes to the extent that they fulfil the eligibility conditions;

(b) organise specific awareness raising activities about the application of this Regulation tailored to the needs of the small-scale providers and users;

(c) where appropriate, establish a dedicated channel for communication with small-scale providers and user and other innovators to provide guidance and respond to queries about the implementation of this Regulation.

2. The specific interests and needs of the small-scale providers shall be taken into account when setting the fees for conformity assessment under Article 43, reducing those fees proportionately to their size and market size.

TITLE VI

GOVERNANCE

CHAPTER 1

EUROPEAN ARTIFICIAL INTELLIGENCE BOARD

Article 56
Establishment of the European Artificial Intelligence Board

1. A 'European Artificial Intelligence Board' (the 'Board') is established.

2. The Board shall provide advice and assistance to the Commission in order to:

   (a) contribute to the effective cooperation of the national supervisory authorities and the Commission with regard to matters covered by this Regulation;

   (b) coordinate and contribute to guidance and analysis by the Commission and the national supervisory authorities and other competent authorities on emerging issues across the internal market with regard to matters covered by this Regulation;

   (c) assist the national supervisory authorities and the Commission in ensuring the consistent application of this Regulation.

Article 57
Structure of the Board

1. The Board shall be composed of the national supervisory authorities, who shall be represented by the head or equivalent high-level official of that authority, and the European Data Protection Supervisor. Other national authorities may be invited to the meetings, where the issues discussed are of relevance for them.

2. The Board shall adopt its rules of procedure by a simple majority of its members, following the consent of the Commission. The rules of procedure shall also contain the operational aspects related to the execution of the Board's tasks as listed in Article 58. The Board may establish sub-groups as appropriate for the purpose of examining specific questions.

3. The Board shall be chaired by the Commission. The Commission shall convene the meetings and prepare the agenda in accordance with the tasks of the Board pursuant to this Regulation and with its rules of procedure. The Commission shall provide administrative and analytical support for the activities of the Board pursuant to this Regulation.

4. The Board may invite external experts and observers to attend its meetings and may hold exchanges with interested third parties to inform its activities to an appropriate extent. To that end the Commission may facilitate exchanges between the Board and other Union bodies, offices, agencies and advisory groups.

Article 58
Tasks of the Board

When providing advice and assistance to the Commission in the context of Article 56(2), the Board shall in particular:

(a) collect and share expertise and best practices among Member States;

(b) contribute to uniform administrative practices in the Member States, including for the functioning of regulatory sandboxes referred to in Article 53;

(c) issue opinions, recommendations or written contributions on matters related to the implementation of this Regulation, in particular

  (i) on technical specifications or existing standards regarding the requirements set out in Title III, Chapter 2,

  (ii) on the use of harmonised standards or common specifications referred to in Articles 40 and 41,

  (iii) on the preparation of guidance documents, including the guidelines concerning the setting of administrative fines referred to in Article 71.

CHAPTER 2

NATIONAL COMPETENT AUTHORITIES

Article 59
Designation of national competent authorities

1. National competent authorities shall be established or designated by each Member State for the purpose of ensuring the application and implementation of this Regulation. National competent authorities shall be organised so as to safeguard the objectivity and impartiality of their activities and tasks.

2. Each Member State shall designate a national supervisory authority among the national competent authorities. The national supervisory authority shall act as notifying authority and market surveillance authority unless a Member State has organisational and administrative reasons to designate more than one authority.

3. Member States shall inform the Commission of their designation or designations and, where applicable, the reasons for designating more than one authority.

4. Member States shall ensure that national competent authorities are provided with adequate financial and human resources to fulfil their tasks under this Regulation. In particular, national competent authorities shall have a sufficient number of personnel permanently available whose competences and expertise shall include an in-depth understanding of artificial intelligence technologies, data and data computing, fundamental rights, health and safety risks and knowledge of existing standards and legal requirements.

5. Member States shall report to the Commission on an annual basis on the status of the financial and human resources of the national competent authorities with an assessment of their adequacy. The Commission shall transmit that information to the Board for discussion and possible recommendations.

6. The Commission shall facilitate the exchange of experience between national competent authorities.

7. National competent authorities may provide guidance and advice on the implementation of this Regulation, including to small-scale providers. Whenever national competent authorities intend to provide guidance and advice with regard to an AI system in areas covered by other Union legislation, the competent national authorities under that Union legislation shall be consulted, as appropriate. Member States may also establish one central contact point for communication with operators.

8. When Union institutions, agencies and bodies fall within the scope of this Regulation, the European Data Protection Supervisor shall act as the competent authority for their supervision.

TITLE VII

EU DATABASE FOR STAND-ALONE HIGH-RISK AI SYSTEMS

Article 60
EU database for stand-alone high-risk AI systems

1. The Commission shall, in collaboration with the Member States, set up and maintain a EU database containing information referred to in paragraph 2 concerning high-risk AI systems referred to in Article 6(2) which are registered in accordance with Article 51.

2. The data listed in Annex VIII shall be entered into the EU database by the providers. The Commission shall provide them with technical and administrative support.

3. Information contained in the EU database shall be accessible to the public.

4. The EU database shall contain personal data only insofar as necessary for collecting and processing information in accordance with this Regulation. That information shall include the names and contact details of natural persons who are responsible for registering the system and have the legal authority to represent the provider.

5. The Commission shall be the controller of the EU database. It shall also ensure to providers adequate technical and administrative support.

TITLE VIII

POST-MARKET MONITORING, INFORMATION SHARING, MARKET SURVEILLANCE

CHAPTER 1

POST-MARKET MONITORING

Article 61
Post-market monitoring by providers and post-market monitoring plan for high-risk AI systems

1. Providers shall establish and document a post-market monitoring system in a manner that is proportionate to the nature of the artificial intelligence technologies and the risks of the high-risk AI system.

2. The post-market monitoring system shall actively and systematically collect, document and analyse relevant data provided by users or collected through other sources on the performance of high-risk AI systems throughout their lifetime, and allow the provider to evaluate the continuous compliance of AI systems with the requirements set out in Title III, Chapter 2.

3. The post-market monitoring system shall be based on a post-market monitoring plan. The post-market monitoring plan shall be part of the technical documentation referred to in Annex IV. The Commission shall adopt an implementing act laying down detailed provisions establishing a template for the post-market monitoring plan and the list of elements to be included in the plan.

4. For high-risk AI systems covered by the legal acts referred to in Annex II, where a post-market monitoring system and plan is already established under that legislation, the elements described in paragraphs 1, 2 and 3 shall be integrated into that system and plan as appropriate.

The first subparagraph shall also apply to high-risk AI systems referred to in point 5(b) of Annex III placed on the market or put into service by credit institutions regulated by Directive 2013/36/EU.

CHAPTER 2

SHARING OF INFORMATION ON INCIDENTS AND MALFUNCTIONING

Article 62
Reporting of serious incidents and of malfunctioning

1. Providers of high-risk AI systems placed on the Union market shall report any serious incident or any malfunctioning of those systems which constitutes a breach of obligations under Union law intended to protect fundamental rights to the market surveillance authorities of the Member States where that incident or breach occurred.

Such notification shall be made immediately after the provider has established a causal link between the AI system and the incident or malfunctioning or the reasonable likelihood of such a link, and, in any event, not later than 15 days after the providers becomes aware of the serious incident or of the malfunctioning.

2. Upon receiving a notification related to a breach of obligations under Union law intended to protect fundamental rights, the market surveillance authority shall inform the national public authorities or bodies referred to in Article 64(3). The Commission shall develop dedicated guidance to facilitate compliance with the obligations set out in paragraph 1. That guidance shall be issued 12 months after the entry into force of this Regulation, at the latest.

3. For high-risk AI systems referred to in point 5(b) of Annex III which are placed on the market or put into service by providers that are credit institutions regulated by Directive 2013/36/EU and for high-risk AI systems which are safety components of devices, or are themselves devices, covered by Regulation (EU) 2017/745 and Regulation (EU) 2017/746, the notification of serious incidents or malfunctioning shall be limited to those that that constitute a breach of obligations under Union law intended to protect fundamental rights.

CHAPTER 3

ENFORCEMENT

Article 63
Market surveillance and control of AI systems in the Union market

1. Regulation (EU) 2019/1020 shall apply to AI systems covered by this Regulation. However, for the purpose of the effective enforcement of this Regulation:

   (a) any reference to an economic operator under Regulation (EU) 2019/1020 shall be understood as including all operators identified in Title III, Chapter 3 of this Regulation;

   (b) any reference to a product under Regulation (EU) 2019/1020 shall be understood as including all AI systems falling within the scope of this Regulation.

2. The national supervisory authority shall report to the Commission on a regular basis the outcomes of relevant market surveillance activities. The national supervisory authority shall report, without delay, to the Commission and relevant national competition authorities any information identified in the course of market surveillance activities that may be of potential interest for the application of Union law on competition rules.

3. For high-risk AI systems, related to products to which legal acts listed in Annex II, section A apply, the market surveillance authority for the purposes of this Regulation shall be the authority responsible for market surveillance activities designated under those legal acts.

4. For AI systems placed on the market, put into service or used by financial institutions regulated by Union legislation on financial services, the market surveillance authority for the purposes of this Regulation shall be the relevant authority responsible for the financial supervision of those institutions under that legislation.

5. For AI systems listed in point 1(a) in so far as the systems are used for law enforcement purposes, points 6 and 7 of Annex III, Member States shall designate as market surveillance authorities for the purposes of this Regulation either the competent data protection supervisory authorities under Directive (EU) 2016/680, or Regulation 2016/679 or the national competent authorities supervising the activities of the law enforcement, immigration or asylum authorities putting into service or using those systems.

6. Where Union institutions, agencies and bodies fall within the scope of this Regulation, the European Data Protection Supervisor shall act as their market surveillance authority.

7. Member States shall facilitate the coordination between market surveillance authorities designated under this Regulation and other relevant national authorities or bodies which supervise the application of Union harmonisation legislation listed in Annex II or other Union legislation that might be relevant for the high-risk AI systems referred to in Annex III.

Article 64
Access to data and documentation

1. Access to data and documentation in the context of their activities, the market surveillance authorities shall be granted full access to the training, validation and testing datasets used by the provider, including through application programming interfaces ('API') or other appropriate technical means and tools enabling remote access.

2. Where necessary to assess the conformity of the high-risk AI system with the requirements set out in Title III, Chapter 2 and upon a reasoned request, the market surveillance authorities shall be granted access to the source code of the AI system.

3. National public authorities or bodies which supervise or enforce the respect of obligations under Union law protecting fundamental rights in relation to the use of high-risk AI systems referred to in Annex III shall have the power to request and access any documentation created or maintained under this Regulation when access to that documentation is necessary for the fulfilment of the competences under their mandate within the limits of their jurisdiction. The relevant public authority or body shall inform the market surveillance authority of the Member State concerned of any such request.

4. By 3 months after the entering into force of this Regulation, each Member State shall identify the public authorities or bodies referred to in paragraph 3 and make a list publicly available on the website of the national supervisory authority. Member States shall notify the list to the Commission and all other Member States and keep the list up to date.

5. Where the documentation referred to in paragraph 3 is insufficient to ascertain whether a breach of obligations under Union law intended to protect fundamental rights has occurred, the public authority or body referred to paragraph 3 may make a reasoned request to the market surveillance authority to organise testing of the high-risk AI system through technical means. The market surveillance authority shall organise the testing with the close involvement of the requesting public authority or body within reasonable time following the request.

6. Any information and documentation obtained by the national public authorities or bodies referred to in paragraph 3 pursuant to the provisions of this Article shall be treated in compliance with the confidentiality obligations set out in Article 70.

Article 65
Procedure for dealing with AI systems presenting a risk at national level

1. AI systems presenting a risk shall be understood as a product presenting a risk defined in Article 3, point 19 of Regulation (EU) 2019/1020 insofar as risks to the health or safety or to the protection of fundamental rights of persons are concerned.

2. Where the market surveillance authority of a Member State has sufficient reasons to consider that an AI system presents a risk as referred to in paragraph 1, they shall carry out an evaluation of the AI system concerned in respect of its compliance with all the requirements and obligations laid down in this Regulation. When risks to the protection of fundamental rights are present, the market surveillance authority shall also inform the relevant national public authorities or bodies referred to in Article 64(3). The relevant operators shall cooperate as necessary with the market

surveillance authorities and the other national public authorities or bodies referred to in Article 64(3).

Where, in the course of that evaluation, the market surveillance authority finds that the AI system does not comply with the requirements and obligations laid down in this Regulation, it shall without delay require the relevant operator to take all appropriate corrective actions to bring the AI system into compliance, to withdraw the AI system from the market, or to recall it within a reasonable period, commensurate with the nature of the risk, as it may prescribe.

The market surveillance authority shall inform the relevant notified body accordingly. Article 18 of Regulation (EU) 2019/1020 shall apply to the measures referred to in the second subparagraph.

3. Where the market surveillance authority considers that non-compliance is not restricted to its national territory, it shall inform the Commission and the other Member States of the results of the evaluation and of the actions which it has required the operator to take.

4. The operator shall ensure that all appropriate corrective action is taken in respect of all the AI systems concerned that it has made available on the market throughout the Union.

5. Where the operator of an AI system does not take adequate corrective action within the period referred to in paragraph 2, the market surveillance authority shall take all appropriate provisional measures to prohibit or restrict the AI system's being made available on its national market, to withdraw the product from that market or to recall it. That authority shall inform the Commission and the other Member States, without delay, of those measures.

6. The information referred to in paragraph 5 shall include all available details, in particular the data necessary for the identification of the non-compliant AI system, the origin of the AI system, the nature of the non-compliance alleged and the risk involved, the nature and duration of the national measures taken and the arguments put forward by the relevant operator. In particular, the market surveillance authorities shall indicate whether the non-compliance is due to one or more of the following:

   (a) a failure of the AI system to meet requirements set out in Title III, Chapter 2;

   (b) shortcomings in the harmonised standards or common specifications referred to in Articles 40 and 41 conferring a presumption of conformity.

7. The market surveillance authorities of the Member States other than the market surveillance authority of the Member State initiating the procedure shall without delay inform the Commission and the other Member States of any measures adopted and of any additional information at their disposal relating to the non-compliance of the AI system concerned, and, in the event of disagreement with the notified national measure, of their objections.

8. Where, within three months of receipt of the information referred to in paragraph 5, no objection has been raised by either a Member State or the Commission in respect of a provisional measure taken by a Member State, that measure shall be deemed justified. This is without prejudice to the procedural rights of the concerned operator in accordance with Article 18 of Regulation (EU) 2019/1020.

9. The market surveillance authorities of all Member States shall ensure that appropriate restrictive measures are taken in respect of the product concerned, such as withdrawal of the product from their market, without delay.

Article 66
Union safeguard procedure

1. Where, within three months of receipt of the notification referred to in Article 65(5), objections are raised by a Member State against a measure taken by another Member State, or where the Commission considers the measure to be contrary to Union law, the Commission shall without delay enter into consultation with the relevant Member State and operator or operators and shall evaluate the national measure. On the basis of the results of that evaluation, the Commission shall decide whether the national measure is justified or not within 9 months from the notification referred to in Article 65(5) and notify such decision to the Member State concerned.

2. If the national measure is considered justified, all Member States shall take the measures necessary to ensure that the non-compliant AI system is withdrawn from their market, and shall inform the Commission accordingly. If the national measure is considered unjustified, the Member State concerned shall withdraw the measure.

3. Where the national measure is considered justified and the non-compliance of the AI system is attributed to shortcomings in the harmonised standards or common specifications referred to in Articles 40 and 41 of this Regulation, the Commission shall apply the procedure provided for in Article 11 of Regulation (EU) No 1025/2012.

Article 67
Compliant AI systems which present a risk

1. Where, having performed an evaluation under Article 65, the market surveillance authority of a Member State finds that although an AI system is in compliance with this Regulation, it presents a risk to the health or safety of persons, to the compliance with obligations under Union or national law intended to protect fundamental rights or to other aspects of public interest protection, it shall require the relevant operator to take all appropriate measures to ensure that the AI system concerned, when placed on the market or put into service, no longer presents that risk, to withdraw the AI system from the market or to recall it within a reasonable period, commensurate with the nature of the risk, as it may prescribe.

2. The provider or other relevant operators shall ensure that corrective action is taken in respect of all the AI systems concerned that they have made available on the market throughout the Union within the timeline prescribed by the market surveillance authority of the Member State referred to in paragraph 1.

3. The Member State shall immediately inform the Commission and the other Member States. That information shall include all available details, in particular the data necessary for the identification of the AI system concerned, the origin and the supply chain of the AI system, the nature of the risk involved and the nature and duration of the national measures taken.

4. The Commission shall without delay enter into consultation with the Member States and the relevant operator and shall evaluate the national measures taken. On the basis

of the results of that evaluation, the Commission shall decide whether the measure is justified or not and, where necessary, propose appropriate measures.

5. The Commission shall address its decision to the Member States.

Article 68
Formal non-compliance

1. Where the market surveillance authority of a Member State makes one of the following findings, it shall require the relevant provider to put an end to the non-compliance concerned:

   (a) the conformity marking has been affixed in violation of Article 49;

   (b) the conformity marking has not been affixed;

   (c) the EU declaration of conformity has not been drawn up;

   (d) the EU declaration of conformity has not been drawn up correctly;

   (e) the identification number of the notified body, which is involved in the conformity assessment procedure, where applicable, has not been affixed;

2. Where the non-compliance referred to in paragraph 1 persists, the Member State concerned shall take all appropriate measures to restrict or prohibit the high-risk AI system being made available on the market or ensure that it is recalled or withdrawn from the market.

TITLE IX

CODES OF CONDUCT

Article 69
Codes of conduct

1. The Commission and the Member States shall encourage and facilitate the drawing up of codes of conduct intended to foster the voluntary application to AI systems other than high-risk AI systems of the requirements set out in Title III, Chapter 2 on the basis of technical specifications and solutions that are appropriate means of ensuring compliance with such requirements in light of the intended purpose of the systems.

2. The Commission and the Board shall encourage and facilitate the drawing up of codes of conduct intended to foster the voluntary application to AI systems of requirements related for example to environmental sustainability, accessibility for persons with a disability, stakeholders participation in the design and development of the AI systems and diversity of development teams on the basis of clear objectives and key performance indicators to measure the achievement of those objectives.

3. Codes of conduct may be drawn up by individual providers of AI systems or by organisations representing them or by both, including with the involvement of users and any interested stakeholders and their representative organisations. Codes of conduct may cover one or more AI systems taking into account the similarity of the intended purpose of the relevant systems.

4. The Commission and the Board shall take into account the specific interests and needs of the small-scale providers and start-ups when encouraging and facilitating the drawing up of codes of conduct.

TITLE X

CONFIDENTIALITY AND PENALTIES

Article 70
Confidentiality

1. National competent authorities and notified bodies involved in the application of this Regulation shall respect the confidentiality of information and data obtained in carrying out their tasks and activities in such a manner as to protect, in particular:

   (a) intellectual property rights, and confidential business information or trade secrets of a natural or legal person, including source code, except the cases referred to in Article 5 of Directive 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure apply.

   (b) the effective implementation of this Regulation, in particular for the purpose of inspections, investigations or audits;(c) public and national security interests;

   (c) integrity of criminal or administrative proceedings.

2. Without prejudice to paragraph 1, information exchanged on a confidential basis between the national competent authorities and between national competent authorities and the Commission shall not be disclosed without the prior consultation of the originating national competent authority and the user when high-risk AI systems referred to in points 1, 6 and 7 of Annex III are used by law enforcement, immigration or asylum authorities, when such disclosure would jeopardise public and national security interests.

   When the law enforcement, immigration or asylum authorities are providers of high-risk AI systems referred to in points 1, 6 and 7 of Annex III, the technical documentation referred to in Annex IV shall remain within the premises of those authorities. Those authorities shall ensure that the market surveillance authorities referred to in Article 63(5) and (6), as applicable, can, upon request, immediately access the documentation or obtain a copy thereof. Only staff of the market surveillance authority holding the appropriate level of security clearance shall be allowed to access that documentation or any copy thereof.

3. Paragraphs 1 and 2 shall not affect the rights and obligations of the Commission, Member States and notified bodies with regard to the exchange of information and the dissemination of warnings, nor the obligations of the parties concerned to provide information under criminal law of the Member States.

4. The Commission and Member States may exchange, where necessary, confidential information with regulatory authorities of third countries with which they have concluded bilateral or multilateral confidentiality arrangements guaranteeing an adequate level of confidentiality.

Article 71
Penalties

1. In compliance with the terms and conditions laid down in this Regulation, Member States shall lay down the rules on penalties, including administrative fines, applicable to infringements of this Regulation and shall take all measures necessary to ensure that they are properly and effectively implemented. The penalties provided for shall be effective, proportionate, and dissuasive. They shall take into particular account the interests of small-scale providers and start-up and their economic viability.

2. The Member States shall notify the Commission of those rules and of those measures and shall notify it, without delay, of any subsequent amendment affecting them.

3. The following infringements shall be subject to administrative fines of up to 30 000 000 EUR or, if the offender is company, up to 6 % of its total worldwide annual turnover for the preceding financial year, whichever is higher:

   (a) non-compliance with the prohibition of the artificial intelligence practices referred to in Article 5;

   (b) non-compliance of the AI system with the requirements laid down in Article 10.

4. The non-compliance of the AI system with any requirements or obligations under this Regulation, other than those laid down in Articles 5 and 10, shall be subject to administrative fines of up to 20 000 000 EUR or, if the offender is a company, up to 4 % of its total worldwide annual turnover for the preceding financial year, whichever is higher.

5. The supply of incorrect, incomplete or misleading information to notified bodies and national competent authorities in reply to a request shall be subject to administrative fines of up to 10 000 000 EUR or, if the offender is a company, up to 2 % of its total worldwide annual turnover for the preceding financial year, whichever is higher.

6. When deciding on the amount of the administrative fine in each individual case, all relevant circumstances of the specific situation shall be taken into account and due regard shall be given to the following:

   (a) the nature, gravity and duration of the infringement and of its consequences;

   (b) whether administrative fines have been already applied by other market surveillance authorities to the same operator for the same infringement.

   (c) the size and market share of the operator committing the infringement;

7. Each Member State shall lay down rules on whether and to what extent administrative fines may be imposed on public authorities and bodies established in that Member State.

8. Depending on the legal system of the Member States, the rules on administrative fines may be applied in such a manner that the fines are imposed by competent national courts of other bodies as applicable in those Member States. The application of such rules in those Member States shall have an equivalent effect.

Article 72
Administrative fines on Union institutions, agencies and bodies

1. The European Data Protection Supervisor may impose administrative fines on Union institutions, agencies and bodies falling within the scope of this Regulation. When deciding whether to impose an administrative fine and deciding on the amount of the administrative fine in each individual case, all relevant circumstances of the specific situation shall be taken into account and due regard shall be given to the following:
   (a) the nature, gravity and duration of the infringement and of its consequences;
   (b) the cooperation with the European Data Protection Supervisor in order to remedy the infringement and mitigate the possible adverse effects of the infringement, including compliance with any of the measures previously ordered by the European Data Protection Supervisor against the Union institution or agency or body concerned with regard to the same subject matter;
   (c) any similar previous infringements by the Union institution, agency or body;
2. The following infringements shall be subject to administrative fines of up to 500 000 EUR:
   (a) non-compliance with the prohibition of the artificial intelligence practices referred to in Article 5;
   (b) non-compliance of the AI system with the requirements laid down in Article 10.
3. The non-compliance of the AI system with any requirements or obligations under this Regulation, other than those laid down in Articles 5 and 10, shall be subject to administrative fines of up to 250 000 EUR.
4. Before taking decisions pursuant to this Article, the European Data Protection Supervisor shall give the Union institution, agency or body which is the subject of the proceedings conducted by the European Data Protection Supervisor the opportunity of being heard on the matter regarding the possible infringement. The European Data Protection Supervisor shall base his or her decisions only on elements and circumstances on which the parties concerned have been able to comment. Complainants, if any, shall be associated closely with the proceedings.
5. The rights of defense of the parties concerned shall be fully respected in the proceedings. They shall be entitled to have access to the European Data Protection Supervisor's file, subject to the legitimate interest of individuals or undertakings in the protection of their personal data or business secrets.
6. Funds collected by imposition of fines in this Article shall be the income of the general budget of the Union.

TITLE XI

DELEGATION OF POWER AND COMMITTEE PROCEDURE

Article 73
Exercise of the delegation

1. The power to adopt delegated acts is conferred on the Commission subject to the conditions laid down in this Article.

2. The delegation of power referred to in Article 4, Article 7(1), Article 11(3), Article 43(5) and (6) and Article 48(5) shall be conferred on the Commission for an indeterminate period of time from [*entering into force of the Regulation*].

3. The delegation of power referred to in Article 4, Article 7(1), Article 11(3), Article 43(5) and (6) and Article 48(5) may be revoked at any time by the European Parliament or by the Council. A decision of revocation shall put an end to the delegation of power specified in that decision. It shall take effect the day following that of its publication in the *Official Journal of the European Union* or at a later date specified therein. It shall not affect the validity of any delegated acts already in force.

4. As soon as it adopts a delegated act, the Commission shall notify it simultaneously to the European Parliament and to the Council.

5. Any delegated act adopted pursuant to Article 4, Article 7(1), Article 11(3), Article 43(5) and (6) and Article 48(5) shall enter into force only if no objection has been expressed by either the European Parliament or the Council within a period of three months of notification of that act to the European Parliament and the Council or if, before the expiry of that period, the European Parliament and the Council have both informed the Commission that they will not object. That period shall be extended by three months at the initiative of the European Parliament or of the Council.

Article 74
Committee procedure

1. The Commission shall be assisted by a committee. That committee shall be a committee within the meaning of Regulation (EU) No 182/2011.

2. Where reference is made to this paragraph, Article 5 of Regulation (EU) No 182/2011 shall apply.

TITLE XII

FINAL PROVISIONS

Article 75
Amendment to Regulation (EC) No 300/2008

In Article 4(3) of Regulation (EC) No 300/2008, the following subparagraph is added:

"When adopting detailed measures related to technical specifications and procedures for approval and use of security equipment concerning Artificial Intelligence systems in the meaning of Regulation (EU) YYY/XX [on Artificial Intelligence] of the European Parliament and of the Council*, the requirements set out in Chapter 2, Title III of that Regulation shall be taken into account."

_____

* Regulation (EU) YYY/XX [on Artificial Intelligence] (OJ …)."

Article 76
Amendment to Regulation (EU) No 167/2013

In Article 17(5) of Regulation (EU) No 167/2013, the following subparagraph is added:

"When adopting delegated acts pursuant to the first subparagraph concerning artificial intelligence systems which are safety components in the meaning of Regulation (EU) YYY/XX [on Artificial Intelligence] of the European Parliament and of the Council*, the requirements set out in Title III, Chapter 2 of that Regulation shall be taken into account.

_____

* Regulation (EU) YYY/XX [on Artificial Intelligence] (OJ …)."

Article 77
Amendment to Regulation (EU) No 168/2013

In Article 22(5) of Regulation (EU) No 168/2013, the following subparagraph is added:

"When adopting delegated acts pursuant to the first subparagraph concerning Artificial Intelligence systems which are safety components in the meaning of Regulation (EU) YYY/XX on [Artificial Intelligence] of the European Parliament and of the Council*, the requirements set out in Title III, Chapter 2 of that Regulation shall be taken into account.

_____

* Regulation (EU) YYY/XX [on Artificial Intelligence] (OJ …)."

Article 78
Amendment to Directive 2014/90/EU

In Article 8 of Directive 2014/90/EU, the following paragraph is added:

"4. For Artificial Intelligence systems which are safety components in the meaning of Regulation (EU) YYY/XX [on Artificial Intelligence] of the European Parliament and of the Council*, when carrying out its activities pursuant to paragraph 1 and when adopting technical specifications and testing standards in accordance with paragraphs 2 and 3, the Commission shall take into account the requirements set out in Title III, Chapter 2 of that Regulation.

_____

* Regulation (EU) YYY/XX [on Artificial Intelligence] (OJ …).".

Article 79
Amendment to Directive (EU) 2016/797

In Article 5 of Directive (EU) 2016/797, the following paragraph is added:

"12. When adopting delegated acts pursuant to paragraph 1 and implementing acts pursuant to paragraph 11 concerning Artificial Intelligence systems which are safety components in the meaning of Regulation (EU) YYY/XX [on Artificial Intelligence] of the European Parliament and of the Council*, the requirements set out in Title III, Chapter 2 of that Regulation shall be taken into account.

_____

* Regulation (EU) YYY/XX [on Artificial Intelligence] (OJ …).".

Article 80
Amendment to Regulation (EU) 2018/858

In Article 5 of Regulation (EU) 2018/858 the following paragraph is added:

"4. When adopting delegated acts pursuant to paragraph 3 concerning Artificial Intelligence systems which are safety components in the meaning of Regulation (EU) YYY/XX [on Artificial Intelligence] of the European Parliament and of the Council *, the requirements set out in Title III, Chapter 2 of that Regulation shall be taken into account.

_____

* Regulation (EU) YYY/XX [on Artificial Intelligence] (OJ …).".

Article 81
Amendment to Regulation (EU) 2018/1139

Regulation (EU) 2018/1139 is amended as follows:

(1) In Article 17, the following paragraph is added:

"3. Without prejudice to paragraph 2, when adopting implementing acts pursuant to paragraph 1 concerning Artificial Intelligence systems which are safety components in the meaning of Regulation (EU) YYY/XX [*on Artificial Intelligence*] of the European Parliament and of the Council*, the requirements set out in Title III, Chapter 2 of that Regulation shall be taken into account.

_____

* Regulation (EU) YYY/XX [on Artificial Intelligence] (OJ …)."

(2) In Article 19, the following paragraph is added:

"4. When adopting delegated acts pursuant to paragraphs 1 and 2 concerning Artificial Intelligence systems which are safety components in the meaning of Regulation (EU) YYY/XX [on Artificial Intelligence], the requirements set out in Title III, Chapter 2 of that Regulation shall be taken into account."

(3) In Article 43, the following paragraph is added:

"4. When adopting implementing acts pursuant to paragraph 1 concerning Artificial Intelligence systems which are safety components in the meaning of Regulation (EU) YYY/XX [on Artificial Intelligence], the requirements set out in Title III, Chapter 2 of that Regulation shall be taken into account."

(4) In Article 47, the following paragraph is added:

"3. When adopting delegated acts pursuant to paragraphs 1 and 2 concerning Artificial Intelligence systems which are safety components in the meaning of Regulation (EU) YYY/XX [on Artificial Intelligence], the requirements set out in Title III, Chapter 2 of that Regulation shall be taken into account."

(5) In Article 57, the following paragraph is added:

"When adopting those implementing acts concerning Artificial Intelligence systems which are safety components in the meaning of Regulation (EU) YYY/XX [on Artificial Intelligence], the requirements set out in Title III, Chapter 2 of that Regulation shall be taken into account."

(6) In Article 58, the following paragraph is added:

"3. When adopting delegated acts pursuant to paragraphs 1 and 2 concerning Artificial Intelligence systems which are safety components in the meaning of Regulation (EU) YYY/XX [on Artificial Intelligence] , the requirements set out in Title III, Chapter 2 of that Regulation shall be taken into account.".

Article 82
Amendment to Regulation (EU) 2019/2144

In Article 11 of Regulation (EU) 2019/2144, the following paragraph is added:

"3. When adopting the implementing acts pursuant to paragraph 2, concerning artificial intelligence systems which are safety components in the meaning of Regulation (EU) YYY/XX [on Artificial Intelligence] of the European Parliament and of the Council*, the requirements set out in Title III, Chapter 2 of that Regulation shall be taken into account.

_____

* Regulation (EU) YYY/XX [on Artificial Intelligence] (OJ …).".

Article 83
AI systems already placed on the market or put into service

1. This Regulation shall not apply to the AI systems which are components of the large-scale IT systems established by the legal acts listed in Annex IX that have been placed on the market or put into service before *[12 months after the date of application of this Regulation referred to in Article 85(2)]*, unless the replacement or amendment of those legal acts leads to a significant change in the design or intended purpose of the AI system or AI systems concerned.

   The requirements laid down in this Regulation shall be taken into account, where applicable, in the evaluation of each large-scale IT systems established by the legal acts listed in Annex IX to be undertaken as provided for in those respective acts.

2. This Regulation shall apply to the high-risk AI systems, other than the ones referred to in paragraph 1, that have been placed on the market or put into service before [*date of application of this Regulation referred to in Article 85(2)*], only if, from that date, those systems are subject to significant changes in their design or intended purpose.

Article 84
Evaluation and review

1. The Commission shall assess the need for amendment of the list in Annex III once a year following the entry into force of this Regulation.

2. By [*three years after the date of application of this Regulation referred to in Article 85(2)*] and every four years thereafter, the Commission shall submit a report on the evaluation and review of this Regulation to the European Parliament and to the Council. The reports shall be made public.

3. The reports referred to in paragraph 2 shall devote specific attention to the following:

   (a) the status of the financial and human resources of the national competent authorities in order to effectively perform the tasks assigned to them under this Regulation;

(b) the state of penalties, and notably administrative fines as referred to in Article 71(1), applied by Member States to infringements of the provisions of this Regulation.

4. Within [*three years after the date of application of this Regulation referred to in Article 85(2)*] and every four years thereafter, the Commission shall evaluate the impact and effectiveness of codes of conduct to foster the application of the requirements set out in Title III, Chapter 2 and possibly other additional requirements for AI systems other than high-risk AI systems.

5. For the purpose of paragraphs 1 to 4 the Board, the Member States and national competent authorities shall provide the Commission with information on its request.

6. In carrying out the evaluations and reviews referred to in paragraphs 1 to 4 the Commission shall take into account the positions and findings of the Board, of the European Parliament, of the Council, and of other relevant bodies or sources.

7. The Commission shall, if necessary, submit appropriate proposals to amend this Regulation, in particular taking into account developments in technology and in the light of the state of progress in the information society.

Article 85
Entry into force and application

1. This Regulation shall enter into force on the twentieth day following that of its publication in the *Official Journal of the European Union*.

2. This Regulation shall apply from [24 months following the entering into force of the Regulation].

3. By way of derogation from paragraph 2:

(a) Title III, Chapter 4 and Title VI shall apply from [three months following the entry into force of this Regulation];

(b) Article 71 shall apply from [twelve months following the entry into force of this Regulation].

This Regulation shall be binding in its entirety and directly applicable in all Member States.

Done at Brussels,

For the European Parliament	*For the Council*
The President	*The President*

LEGISLATIVE FINANCIAL STATEMENT

1. FRAMEWORK OF THE PROPOSAL/INITIATIVE

1.1. Title of the proposal/initiative

1.2. Policy area(s) concerned

1.3. The proposal/initiative relates to:

1.4. Objective(s)

1.4.1. General objective(s)

1.4.2. Specific objective(s)

1.4.3. Expected result(s) and impact

1.4.4. Indicators of performance

1.5. Grounds for the proposal/initiative

1.5.1. Requirement(s) to be met in the short or long term including a detailed timeline for roll-out of the implementation of the initiative

1.5.2. Added value of Union involvement (it may result from different factors, e.g. coordination gains, legal certainty, greater effectiveness or complementarities). For the purposes of this point 'added value of Union involvement' is the value resulting from Union intervention which is additional to the value that would have been otherwise created by Member States alone

1.5.3. Lessons learned from similar experiences in the past

1.5.4. Compatibility with the Multiannual Financial Framework and possible synergies with other appropriate instruments

1.5.5 Assessment of the different available financing options, including scope for redeployment

1.6. Duration and financial impact of the proposal/initiative

1.7. Management mode(s) planned

2. MANAGEMENT MEASURES

2.1. Monitoring and reporting rules

2.2. Management and control system

2.2.1. Justification of the management mode(s), the funding implementation mechanism(s), the payment modalities and the control strategy proposed

2.2.2. Information concerning the risks identified and the internal control system(s) set up to mitigate them

2.2.3. Estimation and justification of the cost-effectiveness of the controls (ratio of "control costs ÷ value of the related funds managed"), and assessment of the expected levels of risk of error (at payment & at closure)

2.3. Measures to prevent fraud and irregularities

3. ESTIMATED FINANCIAL IMPACT OF THE PROPOSAL/INITIATIVE

3.1. Heading(s) of the multiannual financial framework and expenditure budget line(s) affected

3.2. Estimated financial impact of the proposal on appropriations

3.2.1. Summary of estimated impact on operational appropriations

3.2.2. Estimated output funded with operational appropriations

3.2.3. Summary of estimated impact on administrative appropriations

3.2.4. Compatibility with the current multiannual financial framework

3.2.5. Third-party contributions

3.3. Estimated impact on revenue

LEGISLATIVE FINANCIAL STATEMENT

1. FRAMEWORK OF THE PROPOSAL/INITIATIVE

1.1. Title of the proposal/initiative

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

1.2. Policy area(s) concerned

Communications Networks, Content and Technology;

Internal Market, Industry, Entrepreneurship and SMEs;

The budgetary impact concerns the new tasks entrusted with the Commission, including the support to the EU AI Board;

Activity: Shaping Europe's digital future.

1.3. The proposal/initiative relates to:

X **a new action**

☐ **a new action following a pilot project/preparatory action**[64]

☐ **the extension of an existing action**

☐ **an action redirected towards a new action**

1.4. Objective(s)

1.4.1. General objective(s)

The general objective of the intervention is to ensure the proper functioning of the single market by creating the conditions for the development and use of trustworthy artificial intelligence in the Union.

1.4.2. Specific objective(s)

Specific objective No 1

To set requirements specific to AI systems and obligations on all value chain participants in order to ensure that AI systems placed on the market and used are safe and respect existing law on fundamental rights and Union values;

Specific objective No 2

To ensure legal certainty to facilitate investment and innovation in AI by making it clear what essential requirements, obligations, as well as conformity and compliance procedures must be followed to place or use an AI system in the Union market;

Specific objective No 3

To enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems by providing new powers, resources and clear rules for relevant authorities on conformity assessment and ex

[64] As referred to in Article 54(2)(a) or (b) of the Financial Regulation

post monitoring procedures and the division of governance and supervision tasks between national and EU levels;

Specific objective No 4

To facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation by taking EU action to set minimum requirement for AI systems to be placed and used in the Union market in compliance with existing law on fundamental rights and safety.

1.4.3. Expected result(s) and impact

Specify the effects which the proposal/initiative should have on the beneficiaries/groups targeted.

AI suppliers should benefit from a minimal but clear set of requirements, creating legal certainty and ensuring access to the entire single market.

AI users should benefit from legal certainty that the high-risk AI systems they buy comply with European laws and values.

Consumers should benefit by reducing the risk of violations of their safety or fundamental rights.

1.4.4. Indicators of performance

Specify the indicators for monitoring implementation of the proposal/initiative.

Indicator 1

Number of serious incidents or AI performances which constitute a serious incident or a breach of fundamental rights obligations (semi-annual) by fields of applications and calculated a) in absolute terms, b) as share of applications deployed and c) as share of citizens concerned.

Indicator 2

a) Total AI investment in the EU (annual)

b) Total AI investment by Member State (annual)

c) Share of companies using AI (annual)

d) Share of SMEs using AI (annual)

a) and b) will be calculated based on official sources and benchmarked against private estimates

c) and d) will be collected by regular company surveys

1.5. Grounds for the proposal/initiative

1.5.1. Requirement(s) to be met in the short or long term including a detailed timeline for roll-out of the implementation of the initiative

The Regulation should be fully applicable one year and a half after its adoption. However, elements of the governance structure should be in place before then. In particular, Member States shall have appointed existing authorities and/or established new authorities performing the tasks set out in the legislation earlier, and the EU AI Board should be set-up and effective. By the time of applicability, the European database of AI systems should be fully operative. In parallel to the adoption process, it is therefore necessary to develop the database, so that its development has come to an end when the regulation enters into force.

1.5.2. Added value of Union involvement (it may result from different factors, e.g. coordination gains, legal certainty, greater effectiveness or complementarities). For the purposes of this point 'added value of Union involvement' is the value resulting from Union intervention which is additional to the value that would have been otherwise created by Member States alone.

An emerging patchy framework of potentially divergent national rules will hamper the seamless provision of AI systems across the EU and is ineffective in ensuring the

safety and protection of fundamental rights and Union values across the different Member States. A common EU legislative action on AI could boost the internal market and has great potential to provide European industry with a competitive edge at the global scene and economies of scale that cannot be achieved by individual Member States alone.

1.5.3. Lessons learned from similar experiences in the past

The E-commerce Directive 2000/31/EC provides the core framework for the functioning of the single market and the supervision of digital services and sets a basic structure for a general cooperation mechanism among Member States, covering in principle all requirements applicable to digital services. The evaluation of the Directive pointed to shortcomings in several aspects of this cooperation mechanism, including important procedural aspects such as the lack of clear timeframes for response from Member States coupled with a general lack of responsiveness to requests from their counterparts. This has led over the years to a lack of trust between Member States in addressing concerns about providers offering digital services cross-border. The evaluation of the Directive showed the need to define a differentiated set of rules and requirements at European level. For this reason, the implementation of the specific obligations laid down in this Regulation would require a specific cooperation mechanism at EU level, with a governance structure ensuring coordination of specific responsible bodies at EU level.

1.5.4. Compatibility with the Multiannual Financial Framework and possible synergies with other appropriate instruments

The Regulation Laying Down Harmonised Rules on Artificial Intelligence and Amending Certain Union Legislative Acts defines a new common framework of requirements applicable to AI systems, which goes well beyond the framework provided by existing legislation. For this reason, a new national and European regulatory and coordination function needs to be established with this proposal.

As regards possible synergies with other appropriate instruments, the role of notifying authorities at national level can be performed by national authorities fulfilling similar functions sunder other EU regulations.

Moreover, by increasing trust in AI and thus encouraging investment in development and adoption of AI, it complements Digital Europe, for which promoting the diffusion of AI is one of five priorities.

1.5.5. Assessment of the different available financing options, including scope for redeployment

The staff will be redeployed. The other costs will be supported from the DEP. envelope, given that the objective of this regulation – ensuring trustworthy AI – contributes directly to one key objective of Digital Europe – accelerating AI development and deployment in Europe.

1.6. Duration and financial impact of the proposal/initiative

☐ **limited duration**

– ☐ in effect from [DD/MM]YYYY to [DD/MM]YYYY

– ☐ Financial impact from YYYY to YYYY for commitment appropriations and from YYYY to YYYY for payment appropriations.

X **unlimited duration**

– Implementation with a start-up period from one/two (tbc) year,

– followed by full-scale operation.

1.7. Management mode(s) planned[65]

X **Direct management** by the Commission

– ☐ by its departments, including by its staff in the Union delegations;

– ☐ by the executive agencies

☐ **Shared management** with the Member States

☐ **Indirect management** by entrusting budget implementation tasks to:

– ☐ third countries or the bodies they have designated;

– ☐ international organisations and their agencies (to be specified);

– ☐ the EIB and the European Investment Fund;

– ☐ bodies referred to in Articles 70 and 71 of the Financial Regulation;

– ☐ public law bodies;

– ☐ bodies governed by private law with a public service mission to the extent that they provide adequate financial guarantees;

– ☐ bodies governed by the private law of a Member State that are entrusted with the implementation of a public-private partnership and that provide adequate financial guarantees;

– ☐ persons entrusted with the implementation of specific actions in the CFSP pursuant to Title V of the TEU, and identified in the relevant basic act.

– *If more than one management mode is indicated, please provide details in the 'Comments' section.*

Comments

[65] Details of management modes and references to the Financial Regulation may be found on the BudgWeb site: http://www.cc.cec/budg/man/budgmanag/budgmanag_en.html

2. MANAGEMENT MEASURES

2.1. Monitoring and reporting rules

Specify frequency and conditions.

The Regulation will be reviewed and evaluated five years from the entry into force of the regulation. The Commission will report on the findings of the evaluation to the European Parliament, the Council and the European Economic and Social Committee.

2.2. Management and control system(s)

2.2.1. Justification of the management mode(s), the funding implementation mechanism(s), the payment modalities and the control strategy proposed

The Regulation establishes a new policy with regard to harmonised rules for the provision of artificial intelligence systems in the internal market while ensuring the respect of safety and fundamental rights. These new rules require a consistency mechanism for the cross-border application of the obligations under this Regulation in the form of a new advisory group coordinating the activities of national authorities.

In order to face these new tasks, it is necessary to appropriately resource the Commission's services. The enforcement of the new Regulation is estimated to require 10 FTE à regime (5 FTE for the support to the activities of the Board and 5 FTE for the European Data Protection Supervisor acting as a notifying body for AI systems deployed by a body of the European Union).

2.2.2. Information concerning the risks identified and the internal control system(s) set up to mitigate them

In order to ensure that the members of the Board have the possibility to make informed analysis on the basis of factual evidence, it is foreseen that the Board should be supported by the administrative structure of the Commission and that an expert group be created to provide additional expertise where required.

2.2.3. Estimate and justification of the cost-effectiveness of the controls (ratio of "control costs ÷ value of the related funds managed"), and assessment of the expected levels of risk of error (at payment & at closure)

For the meeting expenditure, given the low value per transaction (e.g. refunding travel costs for a delegate for a meeting), standard control procedures seem sufficient. Regarding the development of the database, contract attribution has a strong internal control system in place in DG CNECT through centralised procurement activities.

2.3. Measures to prevent fraud and irregularities

Specify existing or envisaged prevention and protection measures, e.g. from the Anti-Fraud Strategy.

The existing fraud prevention measures applicable to the Commission will cover the additional appropriations necessary for this Regulation.

3. ESTIMATED FINANCIAL IMPACT OF THE PROPOSAL/INITIATIVE

3.1. Heading(s) of the multiannual financial framework and expenditure budget line(s) affected

- Existing budget lines

In order of multiannual financial framework headings and budget lines.

Heading of multiannual financial framework	Budget line	Type of expenditure	Contribution			
	Number	Diff./Non-diff.[66]	from EFTA countries[67]	from candidate countries[68]	from third countries	within the meaning of Article 21(2)(b) of the Financial Regulation
7	20 02 06 Administrative expenditure	Non-diff.	NO	NO	NO	NO
1	02 04 03 DEP Artificial Intelligence	Diff.	YES	NO	NO	NO
1	02 01 30 01 Support expenditure for the Digital Europe programme	Non-diff.	YES	NO	NO	NO

3.2. Estimated financial impact of the proposal on appropriations

3.2.1. Summary of estimated impact on expenditure on operational appropriations

– ☐ The proposal/initiative does not require the use of operational appropriations

– X The proposal/initiative requires the use of operational appropriations, as explained below:

EUR million (to three decimal places)

[66] Diff. = Differentiated appropriations / Non-diff. = Non-differentiated appropriations.

[67] EFTA: European Free Trade Association.

[68] Candidate countries and, where applicable, potential candidate countries from the Western Balkans.

Heading of multiannual financial framework	1	

DG: CNECT				Year **2022**	Year **2023**	Year **2024**	Year **2025**	Year **2026**	Year **2027**[69]	**TOTAL**	
• Operational appropriations											
Budget line[70] 02 04 03		Commitments	(1a)		1.000						**1.000**
		Payments	(2a)		0.600	0.100	0.100	0.100	0.100		**1.000**
Budget line		Commitments	(1b)								
		Payments	(2b)								
Appropriations of an administrative nature financed from the envelope of specific programmes[71]											
Budget line 02 01 30 01			(3)		0.240	0.240	0.240	0.240	0.240		**1.200**
TOTAL appropriations for DG CNECT		Commitments	=1a+1b +3		**1.240**		**0.240**	**0.240**	**0.240**		**2.200**
		Payments	=2a+2b +3		**0.840**	**0.340**	**0.340**	**0.340**	**0.340**		**2.200**

69 Indicative and dependent on budget availability.

70 According to the official budget nomenclature.

71 Technical and/or administrative assistance and expenditure in support of the implementation of EU programmes and/or actions (former 'BA' lines), indirect research, direct research.

• TOTAL operational appropriations	Commitments	(4)		1.000						**1.000**
	Payments	(5)		0.600	0.100	0.100	0.100	0.100		**1.000**
• TOTAL appropriations of an administrative nature financed from the envelope for specific programmes		(6)		**0.240**	**0.240**	**0.240**	**0.240**	**0.240**		**1.200**
TOTAL appropriations under HEADING 1 of the multiannual financial framework	Commitments	=4+ 6		1.240	0.240	0.240	.0.240	0.240		**2.200**
	Payments	=5+ 6		0.840	0.340	0.340	0.340	0.340		**2.200**

If more than one heading is affected by the proposal / initiative, repeat the section above:

• TOTAL operational appropriations (all operational headings)	Commitments	(4)								
	Payments	(5)								
• TOTAL appropriations of an administrative nature financed from the envelope for specific programmes (all operational headings)		(6)								
TOTAL appropriations under HEADINGS 1 to 6 of the multiannual financial framework (Reference amount)	Commitments	=4+ 6								
	Payments	=5+ 6								

Heading of multiannual financial framework	7	'Administrative expenditure'

This section should be filled in using the 'budget data of an administrative nature' to be firstly introduced in the Annex to the Legislative Financial Statement (Annex V to the internal rules), which is uploaded to DECIDE for interservice consultation purposes.

EUR million (to three decimal places)

		Year **2023**	Year **2024**	Year **2025**	Year **2026**	Year **2027**	**After 2027**[72]	**TOTAL**
DG: CNECT								
• Human resources		0.760	0.760	0.760	0.760	0.760	0.760	**3.800**
• Other administrative expenditure		**0.010**	**0.010**	**0.010**	**0.010**	**0.010**	**0.010**	**0.050**
TOTAL DG CNECT	**Appropriations**	**0.760**	**0.760**	**0.760**	**0.760**	**0.760**	**0.760**	**3.850**
European Data Protection Supervisor								
• Human resources		0.760	0.760	0.760	0.760	0.760	0.760	**3.800**
• Other administrative expenditure								
TOTAL EDPS	**Appropriations**	**0.760**	**0.760**	**0.760**	**0.760**	**0.760**	**0.760**	**3.800**
TOTAL appropriations under HEADING 7 of the multiannual financial framework	(Total commitments = Total payments)	1.530	1.530	1.530	1.530	1.530	1.530	**7.650**

EUR million (to three decimal places)

		Year **2022**	Year **2023**	Year **2024**	Year **2025**	Year **2026**	Year **2027**		**TOTAL**
TOTAL appropriations	Commitments		2.770	1.770	1.770	1.770	1.770		**9.850**

[72] All figures in this column are indicative and subject to the continuation of the programmes and availability of appropriations

under HEADINGS 1 to 7 of the multiannual financial framework	Payments		2.370	1.870	1.870	1.870	1.870	**9.850**

3.2.2. Estimated output funded with operational appropriations

Commitment appropriations in EUR million (to three decimal places)

Indicate objectives and outputs ⇩			Year **2022**		Year **2023**		Year **2024**		Year **2025**		Year **2026**		Year **2027**		After **2027**[73]		**TOTAL**	
	OUTPUTS																	
	Type	Average cost	No	Cost	No	Cost	No	Cost	No	Cost	No	Cost	No	Cost	No	Cost	Total No	Total cost
SPECIFIC OBJECTIVE No 1[74]…																		
Database					1	1.000	1		1		1		1		1	0.100	1	1.000
Meetings- Output					10	0.200	10	0.200	10	0.200	10	0.200	10	0.200	10	0.200	50	1.000
Communication activities					2	0.040	2	0.040	2	0.040	2	0.040	2	0.040	2	0.040	10	0.040
Subtotal for specific objective No 1																		
SPECIFIC OBJECTIVE No 2 ...																		
- Output																		
Subtotal for specific objective No 2																		
TOTALS					13	0.240	13	0.240	13	0.240	13	0.240	13	0.240	13	0.100	65	2.200

[73] All figures in this column are indicative and subject to the continuation of the programmes and availability of appropriations

[74] As described in point 1.4.2. 'Specific objective(s)…'

3.2.3. Summary of estimated impact on administrative appropriations

– ☐ The proposal/initiative does not require the use of appropriations of an administrative nature

– X The proposal/initiative requires the use of appropriations of an administrative nature, as explained below:

EUR million (to three decimal places)

	Year **2022**	Year **2023**	Year **2024**	Year **2025**	Year **2026**	Year **2027**	**Yearly after 2027**[75]	**TOTAL**
HEADING 7 of the multiannual financial framework								
Human resources		1.520	1.520	1.520	1.520	1.520	**1.520**	**7.600**
Other administrative expenditure		0.010	0.010	0.010	0.010	0.010	**0.010**	**0.050**
Subtotal HEADING 7 of the multiannual financial framework		1.530	1.530	1.530	1.530	1.530	**1.530**	**7.650**
Outside HEADING 7[76] **of the multiannual financial framework**								
Human resources								
Other expenditure of an administrative nature		0.240	0.240	0.240	0.240	0.240	**0.240**	**1.20**
Subtotal outside HEADING 7 of the multiannual financial framework		0.240	0.240	0.240	0.240	0.240	**0.240**	**1.20**
TOTAL		**1.770**	**1.770**	**1.770**	**1.770**	**1.770**	**1.770**	**8.850**

The appropriations required for human resources and other expenditure of an administrative nature will be met by appropriations from the DG that are already assigned to management of the action and/or have been redeployed within the DG, together if necessary with any additional allocation which may be granted to the managing DG under the annual allocation procedure and in the light of budgetary constraints.

75 All figures in this column are indicative and subject to the continuation of the programmes and availability of appropriations.

76 Technical and/or administrative assistance and expenditure in support of the implementation of EU programmes and/or actions (former 'BA' lines), indirect research, direct research.

3.2.3.1. Estimated requirements of human resources

– ☐ The proposal/initiative does not require the use of human resources.

– X The proposal/initiative requires the use of human resources, as explained below:

Estimate to be expressed in full time equivalent units

.		Year **2023**	Year **2024**	Year **2025**	**2026**	**2027**	**After 2027**[77]	
• Establishment plan posts (officials and temporary staff)								
20 01 02 01 (Headquarters and Commission's Representation Offices)		10	10	10	10	10	10	
20 01 02 03 (Delegations)								
01 01 01 01 (Indirect research)								
01 01 01 11 (Direct research)								
Other budget lines (specify)								
• External staff (in Full Time Equivalent unit: FTE)[78]								
20 02 01 (AC, END, INT from the 'global envelope')								
20 02 03 (AC, AL, END, INT and JPD in the delegations)								
XX 01 xx **yy zz** [79]	- at Headquarters							
	- in Delegations							
01 01 01 02 (AC, END, INT - Indirect research)								
01 01 01 12 (AC, END, INT - Direct research)								
Other budget lines (specify)								
TOTAL		**10**	**10**	**10**	**10**	**10**	**10**	

XX is the policy area or budget title concerned.

The human resources required will be met by staff from the DG who are already assigned to management of the action and/or have been redeployed within the DG, together if necessary with any additional allocation which may be granted to the managing DG under the annual allocation procedure and in the light of budgetary constraints.

EDPS is expected to provide half of the resources required.

Description of tasks to be carried out:

Officials and temporary staff	To prepare a total of 13-16 meetings, draft reports, continue policy work, e.g. regarding future amendments of the list of high-risk AI applications, and maintain relations with Member States' authorities will require four AD FTE and 1 AST FTE. For AI systems developed by the EU institutions, the European Data Protection Supervisor is responsible. Based on past experience, it can be estimated that 5 AD FTE are reuqired to fulfill the EDPS responsibilites under the draft legislation.

[77] All figures in this column are indicative and subject to the continuation of the programmes and availability of appropriations.

[78] AC = Contract Staff; AL = Local Staff; END = Seconded National Expert; INT = agency staff; JPD = Junior Professionals in Delegations.

[79] Sub-ceiling for external staff covered by operational appropriations (former 'BA' lines).

External staff	

3.2.4. Compatibility with the current multiannual financial framework

The proposal/initiative:

– X can be fully financed through redeployment within the relevant heading of the Multiannual Financial Framework (MFF).

No reporgramming is needed.

– ☐ requires use of the unallocated margin under the relevant heading of the MFF and/or use of the special instruments as defined in the MFF Regulation.

Explain what is required, specifying the headings and budget lines concerned, the corresponding amounts, and the instruments proposed to be used.

– ☐ requires a revision of the MFF.

Explain what is required, specifying the headings and budget lines concerned and the corresponding amounts.

3.2.5. Third-party contributions

The proposal/initiative:

– X does not provide for co-financing by third parties

– ☐ provides for the co-financing by third parties estimated below:

Appropriations in EUR million (to three decimal places)

	Year **N**[80]	Year **N+1**	Year **N+2**	Year **N+3**	Enter as many years as necessary to show the duration of the impact (see point 1.6)			Total
Specify the co-financing body								
TOTAL appropriations co-financed								

[80] Year N is the year in which implementation of the proposal/initiative starts. Please replace "N" by the expected first year of implementation (for instance: 2021). The same for the following years.

3.3. Estimated impact on revenue

– ☐ The proposal/initiative has the following financial impact:

– ☐ The proposal/initiative has the following financial impact:

– ☐ on other revenue

– ☐ on other revenue

– Please indicate, if the revenue is assigned to expenditure lines ☐

EUR million (to three decimal places)

Budget revenue line:	Appropriations available for the current financial year	Impact of the proposal/initiative[81]						
		Year **N**	Year **N+1**	Year **N+2**	Year **N+3**	Enter as many years as necessary to show the duration of the impact (see point 1.6)		
Article …….								

For assigned revenue, specify the budget expenditure line(s) affected.

Other remarks (e.g. method/formula used for calculating the impact on revenue or any other information).

[81] As regards traditional own resources (customs duties, sugar levies), the amounts indicated must be net amounts, i.e. gross amounts after deduction of 20 % for collection costs.

Distribution: limited

SHS/IGM-AIETHICS/2021/JUN/3 Rev.2
25 June 2021
Original: English and French

DRAFT TEXT OF THE RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE

Pursuant to the 40 C/Resolution 37, and in accordance with the UNESCO Constitution and the Rules of Procedure concerning recommendations to Member States and international conventions covered by the terms of Article IV, paragraph 4, of the Constitution, the Director-General of UNESCO convened an Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a Recommendation on the Ethics of Artificial Intelligence, and submitted the draft text of the Recommendation to the special committee meeting of technical and legal experts, designated by Member States.

The special committee meeting was held in two phases, the first from 26 to 30 April 2021 and the second from 21 to 25 June 2021. Intersessional consultations were also organized in the period from 1 to 18 June 2021 (12 days). The special committee meeting revised the draft Recommendation and approved the present text for the submission to the General Conference at its 41st session for adoption.

DRAFT TEXT OF THE RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE

PREAMBLE

The General Conference of the United Nations Educational, Scientific and Cultural Organization (UNESCO), meeting in Paris from xx to xx, at its xx session,

Recognizing the profound and dynamic positive and negative impacts of artificial intelligence (AI) on societies, environment, ecosystems and human lives, including the human mind, in part because of the new ways in which its use influences human thinking, interaction and decision-making and affects education, human, social and natural sciences, culture, and communication and information,

Recalling that, by the terms of its Constitution, UNESCO seeks to contribute to peace and security by promoting collaboration among nations through education, the sciences, culture, and communication and information, in order to further universal respect for justice, for the rule of law and for the human rights and fundamental freedoms which are affirmed for the peoples of the world,

Convinced that the Recommendation presented here, as a standard-setting instrument developed through a global approach, based on international law, focusing on human dignity and human rights, as well as gender equality, social and economic justice and development, physical and mental well-being, diversity, interconnectedness, inclusiveness, and environmental and ecosystem protection can guide AI technologies in a responsible direction,

Guided by the purposes and principles of the Charter of the United Nations,

Considering that AI technologies can be of great service to humanity and all countries can benefit from them, but also raise fundamental ethical concerns, for instance regarding the biases they can embed and exacerbate, potentially resulting in discrimination, inequality, digital divides, exclusion and a threat to cultural, social and biological diversity and social or economic divides; the need for transparency and understandability of the workings of algorithms and the data with which they have been trained; and their potential impact on, including but not limited to, human dignity, human rights and fundamental freedoms, gender equality, democracy, social, economic, political and cultural processes, scientific and engineering practices, animal welfare, and the environment and ecosystems,

Also recognizing that AI technologies can deepen existing divides and inequalities in the world, within and between countries, and that justice, trust and fairness must be upheld so that no country and no one should be left behind, either by having fair access to AI technologies and enjoying their benefits or in the protection against their negative implications, while recognizing the different circumstances of different countries and respecting the desire of some people not to take part in all technological developments,

Conscious of the fact that all countries are facing an acceleration in the use of information and communication technologies and AI technologies, as well as an increasing need for media and information literacy, and that the digital economy presents important societal, economic and environmental challenges and opportunities of benefit-sharing, especially for low- and middle-income countries (LMICs), including but not limited to least developed countries (LDCs), landlocked developing countries (LLDCs) and small island developing States (SIDS), requiring the recognition, protection and promotion of endogenous cultures, values and knowledge in order to develop sustainable digital economies,

Further recognizing that AI technologies have the potential to be beneficial to the environment and ecosystems, and in order for those benefits to be realized, potential harms to and negative impacts on the environment and ecosystems should not be ignored but instead addressed,

Noting that addressing risks and ethical concerns should not hamper innovation and development but rather provide new opportunities and stimulate ethically-conducted research and innovation that anchor AI technologies in human rights and fundamental freedoms, values and principles, and moral and ethical reflection,

Also recalling that in November 2019, the General Conference of UNESCO, at its 40th session, adopted 40 C/Resolution 37, by which it mandated the Director-General "to prepare an international standard-setting instrument on the ethics of artificial intelligence (AI) in the form of a recommendation", which is to be submitted to the General Conference at its 41st session in 2021,

Recognizing that the development of AI technologies necessitates a commensurate increase in data, media and information literacy as well as access to independent, pluralistic, trusted sources of information, including as part of efforts to mitigate risks of misinformation, disinformation and hate speech, and harm caused through the misuse of personal data,

Observing that a normative framework for AI technologies and its social implications finds its basis in international and national legal frameworks, human rights and fundamental freedoms, ethics, need for access to data, information and knowledge, the freedom of research and innovation, human and environmental and ecosystem well-being, and connects ethical values and principles to the challenges and opportunities linked to AI technologies, based on common understanding and shared aims,

Also recognizing that ethical values and principles can help develop and implement rights-based policy measures and legal norms, by providing guidance with a view to the fast pace of technological development,

Also convinced that globally accepted ethical standards for AI technologies, in full respect of international law, in particular human rights law, can play a key role in developing AI-related norms across the globe,

Bearing in mind the Universal Declaration of Human Rights (1948), the instruments of the international human rights framework, including the Convention Relating to the Status of Refugees (1951), the Discrimination (Employment and Occupation) Convention (1958), the International Convention on the Elimination of All Forms of Racial Discrimination (1965), the International Covenant on Civil and Political Rights (1966), the International Covenant on Economic, Social and Cultural Rights (1966), the Convention on the Elimination of All Forms of Discrimination against Women (1979), the Convention on the Rights of the Child (1989), and the Convention on the Rights of Persons with Disabilities (2006), the Convention against Discrimination in Education (1960), the Convention on the Protection and Promotion of the Diversity of Cultural Expressions (2005), as well as any other relevant international instruments, recommendations and declarations,

Also noting the United Nations Declaration on the Right to Development (1986); the Declaration on the Responsibilities of the Present Generations Towards Future Generations (1997); the Universal Declaration on Bioethics and Human Rights (2005); the United Nations Declaration on the Rights of Indigenous Peoples (2007); the United Nations General Assembly resolution on the review of the World Summit on the Information Society (A/RES/70/125) (2015); the United Nations General Assembly Resolution on Transforming our world: the 2030 Agenda for Sustainable Development (A/RES/70/1) (2015); the Recommendation Concerning

the Preservation of, and Access to, Documentary Heritage Including in Digital Form (2015); the Declaration of Ethical Principles in relation to Climate Change (2017); the Recommendation on Science and Scientific Researchers (2017); the Internet Universality Indicators (endorsed by UNESCO's International Programme for the Development of Communication in 2018), including the ROAM principles (endorsed by UNESCO's General Conference in 2015); the Human Rights Council's resolution on "The right to privacy in the digital age" (A/HRC/RES/42/15) (2019); and the Human Rights Council's resolution on "New and emerging digital technologies and human rights" (A/HRC/RES/41/11) (2019),

Emphasizing that specific attention must be paid to LMICs, including but not limited to LDCs, LLDCs and SIDS, as they have their own capacity but have been underrepresented in the AI ethics debate, which raises concerns about neglecting local knowledge, cultural pluralism, value systems and the demands of global fairness to deal with the positive and negative impacts of AI technologies,

Also conscious of the many existing national policies, other frameworks and initiatives elaborated by relevant United Nations entities, intergovernmental organizations, including regional organizations, as well as those by the private sector, professional organizations, non-governmental organizations, and the scientific community, related to the ethics and regulation of AI technologies,

Further convinced that AI technologies can bring important benefits, but that achieving them can also amplify tension around innovation, asymmetric access to knowledge and technologies, including the digital and civic literacy deficit that limits the public's ability to engage in topics related to AI, as well as barriers to access to information and gaps in capacity, human and institutional capacities, barriers to access to technological innovation, and a lack of adequate physical and digital infrastructure and regulatory frameworks, including those related to data, all of which need to be addressed,

Underlining that the strengthening of global cooperation and solidarity, including through multilateralism, is needed to facilitate fair access to AI technologies and address the challenges that they bring to diversity and interconnectivity of cultures and ethical systems, to mitigate potential misuse, to realize the full potential that AI can bring, especially in the area of development, and to ensure that national AI strategies are guided by ethical principles,

Taking fully into account that the rapid development of AI technologies challenges their ethical implementation and governance, as well as the respect for and protection of cultural diversity, and has the potential to disrupt local and regional ethical standards and values,

1. ***Adopts*** the present Recommendation on the Ethics of Artificial Intelligence;

2. ***Recommends*** that Member States apply on a voluntary basis the provisions of this Recommendation by taking appropriate steps, including whatever legislative or other measures may be required, in conformity with the constitutional practice and governing structures of each State, to give effect within their jurisdictions to the principles and norms of the Recommendation in conformity with international law, including international human rights law;

3. ***Also recommends*** that Member States engage all stakeholders, including business enterprises, to ensure that they play their respective roles in the implementation of this Recommendation; and bring the Recommendation to the attention of the authorities, bodies, research and academic organizations, institutions and organizations in public, private and civil society sectors involved in AI technologies, so that the development and use of AI technologies are guided by both sound scientific research as well as ethical analysis and evaluation.

I. SCOPE OF APPLICATION

1. This Recommendation addresses ethical issues related to the domain of Artificial Intelligence to the extent that they are within UNESCO's mandate. It approaches AI ethics as a systematic normative reflection, based on a holistic, comprehensive, multicultural and evolving framework of interdependent values, principles and actions that can guide societies in dealing responsibly with the known and unknown impacts of AI technologies on human beings, societies and the environment and ecosystems, and offers them a basis to accept or reject AI technologies. It considers ethics as a dynamic basis for the normative evaluation and guidance of AI technologies, referring to human dignity, well-being and the prevention of harm as a compass and as rooted in the ethics of science and technology.

2. This Recommendation does not have the ambition to provide one single definition of AI, since such a definition would need to change over time, in accordance with technological developments. Rather, its ambition is to address those features of AI systems that are of central ethical relevance. Therefore, this Recommendation approaches AI systems as systems which have the capacity to process data and information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning or control. Three elements have a central place in this approach:

(a) AI systems are information-processing technologies that integrate models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in material and virtual environments. AI systems are designed to operate with varying degrees of autonomy by means of knowledge modelling and representation and by exploiting data and calculating correlations. AI systems may include several methods, such as but not limited to:

(i) machine learning, including deep learning and reinforcement learning;

(ii) machine reasoning, including planning, scheduling, knowledge representation and reasoning, search, and optimization.

AI systems can be used in cyber-physical systems, including the Internet of things, robotic systems, social robotics, and human-computer interfaces, which involve control, perception, the processing of data collected by sensors, and the operation of actuators in the environment in which AI systems work.

(b) Ethical questions regarding AI systems pertain to all stages of the AI system life cycle, understood here to range from research, design and development to deployment and use, including maintenance, operation, trade, financing, monitoring and evaluation, validation, end-of-use, disassembly and termination. In addition, AI actors can be defined as any actor involved in at least one stage of the AI system life cycle, and can refer both to natural and legal persons, such as researchers, programmers, engineers, data scientists, end-users, business enterprises, universities and public and private entities, among others.

(c) AI systems raise new types of ethical issues that include, but are not limited to, their impact on decision-making, employment and labour, social interaction, health care, education, media, access to information, digital divide, personal data and consumer protection, environment, democracy, rule of law, security and policing, dual use, and human rights and fundamental freedoms, including freedom of expression, privacy and non-discrimination. Furthermore, new ethical challenges are created by the potential of AI algorithms to reproduce and reinforce existing biases, and thus to exacerbate already existing forms of discrimination, prejudice

and stereotyping. Some of these issues are related to the capacity of AI systems to perform tasks which previously only living beings could do, and which were in some cases even limited to human beings only. These characteristics give AI systems a profound, new role in human practices and society, as well as in their relationship with the environment and ecosystems, creating a new context for children and young people to grow up in, develop an understanding of the world and themselves, critically understand media and information, and learn to make decisions. In the long term, AI systems could challenge humans' special sense of experience and agency, raising additional concerns about, *inter alia*, human self-understanding, social, cultural and environmental interaction, autonomy, agency, worth and dignity.

3. This Recommendation pays specific attention to the broader ethical implications of AI systems in relation to the central domains of UNESCO: education, science, culture, and communication and information, as explored in the 2019 Preliminary Study on the Ethics of Artificial Intelligence by the UNESCO World Commission on Ethics of Scientific Knowledge and Technology (COMEST):

(a) Education, because living in digitalizing societies requires new educational practices, ethical reflection, critical thinking, responsible design practices and new skills, given the implications for the labour market, employability and civic participation.

(b) Science, in the broadest sense and including all academic fields from the natural sciences and medical sciences to the social sciences and humanities, as AI technologies bring new research capacities and approaches, have implications for our concepts of scientific understanding and explanation, and create a new basis for decision-making.

(c) Cultural identity and diversity, as AI technologies can enrich cultural and creative industries, but can also lead to an increased concentration of supply of cultural content, data, markets and income in the hands of only a few actors, with potential negative implications for the diversity and pluralism of languages, media, cultural expressions, participation and equality.

(d) Communication and information, as AI technologies play an increasingly important role in the processing, structuring and provision of information; the issues of automated journalism and the algorithmic provision of news and moderation and curation of content on social media and search engines are just a few examples raising issues related to access to information, disinformation, misinformation, hate speech, the emergence of new forms of societal narratives, discrimination, freedom of expression, privacy and media and information literacy, among others.

4. This Recommendation is addressed to Member States, both as AI actors and as authorities responsible for developing legal and regulatory frameworks throughout the entire AI system life cycle, and for promoting business responsibility. It also provides ethical guidance to all AI actors, including the public and private sectors, by providing a basis for an ethical impact assessment of AI systems throughout their life cycle.

II. AIMS AND OBJECTIVES

5. This Recommendation aims to provide a basis to make AI systems work for the good of humanity, individuals, societies and the environment and ecosystems, and to prevent harm. It also aims at stimulating the peaceful use of AI systems.

6. In addition to the existing ethical frameworks regarding AI around the world, this Recommendation aims to bring a globally accepted normative instrument that focuses not only on the articulation of values and principles, but also on their practical realization, via concrete policy recommendations, with a strong emphasis on inclusion issues of gender equality and protection of the environment and ecosystems.

7. Because the complexity of the ethical issues surrounding AI necessitates the cooperation of multiple stakeholders across the various levels and sectors of international, regional and national communities, this Recommendation aims to enable stakeholders to take shared responsibility based on a global and intercultural dialogue.

8. The objectives of this Recommendation are:

(a) to provide a universal framework of values, principles and actions to guide States in the formulation of their legislation, policies or other instruments regarding AI, consistent with international law;

(b) to guide the actions of individuals, groups, communities, institutions and private sector companies to ensure the embedding of ethics in all stages of the AI system life cycle;

(c) to protect, promote and respect human rights and fundamental freedoms, human dignity and equality, including gender equality; to safeguard the interests of present and future generations; to preserve the environment, biodiversity and ecosystems; and to respect cultural diversity in all stages of the AI system life cycle;

(d) to foster multi-stakeholder, multidisciplinary and pluralistic dialogue and consensus building about ethical issues relating to AI systems;

(e) to promote equitable access to developments and knowledge in the field of AI and the sharing of benefits, with particular attention to the needs and contributions of LMICs, including LDCs, LLDCs and SIDS.

III. VALUES AND PRINCIPLES

9. The values and principles included below should be respected by all actors in the AI system life cycle, in the first place and, where needed and appropriate, be promoted through amendments to the existing and elaboration of new legislation, regulations and business guidelines. This must comply with international law, including the United Nations Charter and Member States' human rights obligations, and should be in line with internationally agreed social, political, environmental, educational, scientific and economic sustainability objectives, such as the United Nations Sustainable Development Goals (SDGs).

10. Values play a powerful role as motivating ideals in shaping policy measures and legal norms. While the set of values outlined below thus inspires desirable behaviour and represents the foundations of principles, the principles unpack the values underlying them more concretely so that the values can be more easily operationalized in policy statements and actions.

11. While all the values and principles outlined below are desirable per se, in any practical contexts, there may be tensions between these values and principles. In any given situation, a contextual assessment will be necessary to manage potential tensions, taking into account the principle of proportionality and in compliance with human rights and fundamental freedoms. In all cases, any possible limitations on human rights and fundamental freedoms must have a lawful basis, and be reasonable, necessary and proportionate, and consistent with States' obligations under international law. To navigate such scenarios judiciously will typically require

engagement with a broad range of appropriate stakeholders, making use of social dialogue, as well as ethical deliberation, due diligence and impact assessment.

12. The trustworthiness and integrity of the life cycle of AI systems is essential to ensure that AI technologies will work for the good of humanity, individuals, societies and the environment and ecosystems, and embody the values and principles set out in this Recommendation. People should have good reason to trust that AI systems can bring individual and shared benefits, while adequate measures are taken to mitigate risks. An essential requirement for trustworthiness is that, throughout their life cycle, AI systems are subject to thorough monitoring by the relevant stakeholders as appropriate. As trustworthiness is an outcome of the operationalization of the principles in this document, the policy actions proposed in this Recommendation are all directed at promoting trustworthiness in all stages of the AI system life cycle.

III.1 VALUES

Respect, protection and promotion of human rights and fundamental freedoms and human dignity

13. The inviolable and inherent dignity of every human constitutes the foundation for the universal, indivisible, inalienable, interdependent and interrelated system of human rights and fundamental freedoms. Therefore, respect, protection and promotion of human dignity and rights as established by international law, including international human rights law, is essential throughout the life cycle of AI systems. Human dignity relates to the recognition of the intrinsic and equal worth of each individual human being, regardless of race, colour, descent, gender, age, language, religion, political opinion, national origin, ethnic origin, social origin, economic or social condition of birth, or disability and any other grounds.

14. No human being or human community should be harmed or subordinated, whether physically, economically, socially, politically, culturally or mentally during any phase of the life cycle of AI systems. Throughout the life cycle of AI systems, the quality of life of human beings should be enhanced, while the definition of “quality of life” should be left open to individuals or groups, as long as there is no violation or abuse of human rights and fundamental freedoms, or the dignity of humans in terms of this definition.

15. Persons may interact with AI systems throughout their life cycle and receive assistance from them, such as care for vulnerable people or people in vulnerable situations, including but not limited to children, older persons, persons with disabilities or the ill. Within such interactions, persons should never be objectified, nor should their dignity be otherwise undermined, or human rights and fundamental freedoms violated or abused.

16. Human rights and fundamental freedoms must be respected, protected and promoted throughout the life cycle of AI systems. Governments, private sector, civil society, international organizations, technical communities and academia must respect human rights instruments and frameworks in their interventions in the processes surrounding the life cycle of AI systems. New technologies need to provide new means to advocate, defend and exercise human rights and not to infringe them.

Environment and ecosystem flourishing

17. Environmental and ecosystem flourishing should be recognized, protected and promoted through the life cycle of AI systems. Furthermore, environment and ecosystems are the existential necessity for humanity and other living beings to be able to enjoy the benefits of advances in AI.

18. All actors involved in the life cycle of AI systems must comply with applicable international law and domestic legislation, standards and practices, such as precaution, designed for environmental and ecosystem protection and restoration, and sustainable development. They should reduce the environmental impact of AI systems, including but not limited to its carbon footprint, to ensure the minimization of climate change and environmental risk factors, and prevent the unsustainable exploitation, use and transformation of natural resources contributing to the deterioration of the environment and the degradation of ecosystems.

Ensuring diversity and inclusiveness

19. Respect, protection and promotion of diversity and inclusiveness should be ensured throughout the life cycle of AI systems, consistent with international law, including human rights law. This may be done by promoting active participation of all individuals or groups regardless of race, colour, descent, gender, age, language, religion, political opinion, national origin, ethnic origin, social origin, economic or social condition of birth, or disability and any other grounds.

20. The scope of lifestyle choices, beliefs, opinions, expressions or personal experiences, including the optional use of AI systems and the co-design of these architectures should not be restricted during any phase of the life cycle of AI systems.

21. Furthermore, efforts, including international cooperation, should be made to overcome, and never take advantage of, the lack of necessary technological infrastructure, education and skills, as well as legal frameworks, particularly in LMICs, LDCs, LLDCs and SIDS, affecting communities.

Living in peaceful, just and interconnected societies

22. AI actors should play a participative and enabling role to ensure peaceful and just societies, which is based on an interconnected future for the benefit of all, consistent with human rights and fundamental freedoms. The value of living in peaceful and just societies points to the potential of AI systems to contribute throughout their life cycle to the interconnectedness of all living creatures with each other and with the natural environment.

23. The notion of humans being interconnected is based on the knowledge that every human belongs to a greater whole, which thrives when all its constituent parts are enabled to thrive. Living in peaceful, just and interconnected societies requires an organic, immediate, uncalculated bond of solidarity, characterized by a permanent search for peaceful relations, tending towards care for others and the natural environment in the broadest sense of the term.

24. This value demands that peace, inclusiveness and justice, equity and interconnectedness should be promoted throughout the life cycle of AI systems, in so far as the processes of the life cycle of AI systems should not segregate, objectify or undermine freedom and autonomous decision-making as well as the safety of human beings and communities, divide and turn individuals and groups against each other, or threaten the coexistence between humans, other living beings and the natural environment.

III.2 PRINCIPLES

Proportionality and Do No Harm

25. It should be recognized that AI technologies do not necessarily, per se, ensure human and environmental and ecosystem flourishing. Furthermore, none of the processes related to the AI system life cycle shall exceed what is necessary to achieve legitimate aims or objectives

and should be appropriate to the context. In the event of possible occurrence of any harm to human beings, human rights and fundamental freedoms, communities and society at large or the environment and ecosystems, the implementation of procedures for risk assessment and the adoption of measures in order to preclude the occurrence of such harm should be ensured.

26. The choice to use AI systems and which AI method to use should be justified in the following ways: (a) the AI method chosen should be appropriate and proportional to achieve a given legitimate aim; (b) the AI method chosen should not infringe upon the foundational values captured in this document, in particular, its use must not violate or abuse human rights; and (c) the AI method should be appropriate to the context and should be based on rigorous scientific foundations. In scenarios where decisions are understood to have an impact that is irreversible or difficult to reverse or may involve life and death decisions, final human determination should apply. In particular, AI systems should not be used for social scoring or mass surveillance purposes.

Safety and security

27. Unwanted harms (safety risks), as well as vulnerabilities to attack (security risks) should be avoided and should be addressed, prevented and eliminated throughout the life cycle of AI systems to ensure human, environmental and ecosystem safety and security. Safe and secure AI will be enabled by the development of sustainable, privacy-protective data access frameworks that foster better training and validation of AI models utilizing quality data.

Fairness and non-discrimination

28. AI actors should promote social justice and safeguard fairness and non-discrimination of any kind in compliance with international law. This implies an inclusive approach to ensuring that the benefits of AI technologies are available and accessible to all, taking into consideration the specific needs of different age groups, cultural systems, different language groups, persons with disabilities, girls and women, and disadvantaged, marginalized and vulnerable people or people in vulnerable situations. Member States should work to promote inclusive access for all, including local communities, to AI systems with locally relevant content and services, and with respect for multilingualism and cultural diversity. Member States should work to tackle digital divides and ensure inclusive access to and participation in the development of AI. At the national level, Member States should promote equity between rural and urban areas, and among all persons regardless of race, colour, descent, gender, age, language, religion, political opinion, national origin, ethnic origin, social origin, economic or social condition of birth, or disability and any other grounds, in terms of access to and participation in the AI system life cycle. At the international level, the most technologically advanced countries have a responsibility of solidarity with the least advanced to ensure that the benefits of AI technologies are shared such that access to and participation in the AI system life cycle for the latter contributes to a fairer world order with regard to information, communication, culture, education, research and socio-economic and political stability.

29. AI actors should make all reasonable efforts to minimize and avoid reinforcing or perpetuating discriminatory or biased applications and outcomes throughout the life cycle of the AI system to ensure fairness of such systems. Effective remedy should be available against discrimination and biased algorithmic determination.

30. Furthermore, digital and knowledge divides within and between countries need to be addressed throughout an AI system life cycle, including in terms of access and quality of access to technology and data, in accordance with relevant national, regional and international legal frameworks, as well as in terms of connectivity, knowledge and skills and meaningful participation of the affected communities, such that every person is treated equitably.

Sustainability

31. The development of sustainable societies relies on the achievement of a complex set of objectives on a continuum of human, social, cultural, economic and environmental dimensions. The advent of AI technologies can either benefit sustainability objectives or hinder their realization, depending on how they are applied across countries with varying levels of development. The continuous assessment of the human, social, cultural, economic and environmental impact of AI technologies should therefore be carried out with full cognizance of the implications of AI technologies for sustainability as a set of constantly evolving goals across a range of dimensions, such as currently identified in the Sustainable Development Goals (SDGs) of the United Nations.

Right to Privacy, and Data Protection

32. Privacy, a right essential to the protection of human dignity, human autonomy and human agency, must be respected, protected and promoted throughout the life cycle of AI systems. It is important that data for AI systems be collected, used, shared, archived and deleted in ways that are consistent with international law and in line with the values and principles set forth in this Recommendation, while respecting relevant national, regional and international legal frameworks.

33. Adequate data protection frameworks and governance mechanisms should be established in a multi-stakeholder approach at the national or international level, protected by judicial systems, and ensured throughout the life cycle of AI systems. Data protection frameworks and any related mechanisms should take reference from international data protection principles and standards concerning the collection, use and disclosure of personal data and exercise of their rights by data subjects while ensuring a legitimate aim and a valid legal basis for the processing of personal data, including informed consent.

34. Algorithmic systems require adequate privacy impact assessments, which also include societal and ethical considerations of their use and an innovative use of the privacy by design approach. AI actors need to ensure that they are accountable for the design and implementation of AI systems in such a way as to ensure that personal information is protected throughout the life cycle of the AI system.

Human oversight and determination

35. Member States should ensure that it is always possible to attribute ethical and legal responsibility for any stage of the life cycle of AI systems, as well as in cases of remedy related to AI systems, to physical persons or to existing legal entities. Human oversight refers thus not only to individual human oversight, but to inclusive public oversight, as appropriate.

36. It may be the case that sometimes humans would choose to rely on AI systems for reasons of efficacy, but the decision to cede control in limited contexts remains that of humans, as humans can resort to AI systems in decision-making and acting, but an AI system can never replace ultimate human responsibility and accountability. As a rule, life and death decisions should not be ceded to AI systems.

Transparency and explainability

37. The transparency and explainability of AI systems are often essential preconditions to ensure the respect, protection and promotion of human rights, fundamental freedoms and ethical principles. Transparency is necessary for relevant national and international liability regimes to work effectively. A lack of transparency could also undermine the possibility of effectively challenging decisions based on outcomes produced by AI systems

and may thereby infringe the right to a fair trial and effective remedy, and limits the areas in which these systems can be legally used.

38. While efforts need to be made to increase transparency and explainability of AI systems, including those with extra-territorial impact, throughout their life cycle to support democratic governance, the level of transparency and explainability should always be appropriate to the context and impact, as there may be a need to balance between transparency and explainability and other principles such as privacy, safety and security. People should be fully informed when a decision is informed by or is made on the basis of AI algorithms, including when it affects their safety or human rights, and in those circumstances should have the opportunity to request explanatory information from the relevant AI actor or public sector institutions. In addition, individuals should be able to access the reasons for a decision affecting their rights and freedoms, and have the option of making submissions to a designated staff member of the private sector company or public sector institution able to review and correct the decision. AI actors should inform users when a product or service is provided directly or with the assistance of AI systems in a proper and timely manner.

39. From a socio-technical lens, greater transparency contributes to more peaceful, just, democratic and inclusive societies. It allows for public scrutiny that can decrease corruption and discrimination, and can also help detect and prevent negative impacts on human rights. Transparency aims at providing appropriate information to the respective addressees to enable their understanding and foster trust. Specific to the AI system, transparency can enable people to understand how each stage of an AI system is put in place, appropriate to the context and sensitivity of the AI system. It may also include insight into factors that affect a specific prediction or decision, and whether or not appropriate assurances (such as safety or fairness measures) are in place. In cases of serious threats of adverse human rights impacts, transparency may also require the sharing of code or datasets.

40. Explainability refers to making intelligible and providing insight into the outcome of AI systems. The explainability of AI systems also refers to the understandability of the input, output and the functioning of each algorithmic building block and how it contributes to the outcome of the systems. Thus, explainability is closely related to transparency, as outcomes and sub-processes leading to outcomes should aim to be understandable and traceable, appropriate to the context. AI actors should commit to ensuring that the algorithms developed are explainable. In the case of AI applications that impact the end user in a way that is not temporary, easily reversible or otherwise low risk, it should be ensured that the meaningful explanation is provided with any decision that resulted in the action taken in order for the outcome to be considered transparent.

41. Transparency and explainability relate closely to adequate responsibility and accountability measures, as well as to the trustworthiness of AI systems.

Responsibility and accountability

42. AI actors and Member States should respect, protect and promote human rights and fundamental freedoms, and should also promote the protection of the environment and ecosystems, assuming their respective ethical and legal responsibility, in accordance with national and international law, in particular Member States' human rights obligations, and ethical guidance throughout the life cycle of AI systems, including with respect to AI actors within their effective territory and control. The ethical responsibility and liability for the decisions and actions based in any way on an AI system should always ultimately be attributable to AI actors corresponding to their role in the life cycle of the AI system.

43. Appropriate oversight, impact assessment, audit and due diligence mechanisms, including whistle-blowers' protection, should be developed to ensure accountability for AI

systems and their impact throughout their life cycle. Both technical and institutional designs should ensure auditability and traceability of (the working of) AI systems in particular to address any conflicts with human rights norms and standards and threats to environmental and ecosystem well-being.

Awareness and literacy

44. Public awareness and understanding of AI technologies and the value of data should be promoted through open and accessible education, civic engagement, digital skills and AI ethics training, media and information literacy and training led jointly by governments, intergovernmental organizations, civil society, academia, the media, community leaders and the private sector, and considering the existing linguistic, social and cultural diversity, to ensure effective public participation so that all members of society can take informed decisions about their use of AI systems and be protected from undue influence.

45. Learning about the impact of AI systems should include learning about, through and for human rights and fundamental freedoms, meaning that the approach and understanding of AI systems should be grounded by their impact on human rights and access to rights, as well as on the environment and ecosystems.

Multi-stakeholder and adaptive governance and collaboration

46. International law and national sovereignty must be respected in the use of data. That means that States, complying with international law, can regulate the data generated within or passing through their territories, and take measures towards effective regulation of data, including data protection, based on respect for the right to privacy in accordance with international law and other human rights norms and standards.

47. Participation of different stakeholders throughout the AI system life cycle is necessary for inclusive approaches to AI governance, enabling the benefits to be shared by all, and to contribute to sustainable development. Stakeholders include but are not limited to governments, intergovernmental organizations, the technical community, civil society, researchers and academia, media, education, policy-makers, private sector companies, human rights institutions and equality bodies, anti-discrimination monitoring bodies, and groups for youth and children. The adoption of open standards and interoperability to facilitate collaboration should be in place. Measures should be adopted to take into account shifts in technologies, the emergence of new groups of stakeholders, and to allow for meaningful participation by marginalized groups, communities and individuals and, where relevant, in the case of Indigenous Peoples, respect for the self-governance of their data.

IV. AREAS OF POLICY ACTION

48. The policy actions described in the following policy areas operationalize the values and principles set out in this Recommendation. The main action is for Member States to put in place effective measures, including, for example, policy frameworks or mechanisms, and to ensure that other stakeholders, such as private sector companies, academic and research institutions, and civil society adhere to them by, among other actions, encouraging all stakeholders to develop human rights, rule of law, democracy, and ethical impact assessment and due diligence tools in line with guidance including the United Nations Guiding Principles on Business and Human Rights. The process for developing such policies or mechanisms should be inclusive of all stakeholders and should take into account the circumstances and priorities of each Member State. UNESCO can be a partner and support Member States in the development as well as monitoring and evaluation of policy mechanisms.

49. UNESCO recognizes that Member States will be at different stages of readiness to implement this Recommendation, in terms of scientific, technological, economic, educational, legal, regulatory, infrastructural, societal, cultural and other dimensions. It is noted that "readiness" here is a dynamic status. In order to enable the effective implementation of this Recommendation, UNESCO will therefore: (1) develop a readiness assessment methodology to assist interested Member States in identifying their status at specific moments of their readiness trajectory along a continuum of dimensions; and (2) ensure support for interested Member States in terms of developing a UNESCO methodology for Ethical Impact Assessment (EIA) of AI technologies, sharing of best practices, assessment guidelines and other mechanisms and analytical work.

POLICY AREA 1: ETHICAL IMPACT ASSESSMENT

50. Member States should introduce frameworks for impact assessments, such as ethical impact assessment, to identify and assess benefits, concerns and risks of AI systems, as well as appropriate risk prevention, mitigation and monitoring measures, among other assurance mechanisms. Such impact assessments should identify impacts on human rights and fundamental freedoms, in particular but not limited to the rights of marginalized and vulnerable people or people in vulnerable situations, labour rights, the environment and ecosystems and ethical and social implications, and facilitate citizen participation in line with the values and principles set forth in this Recommendation.

51. Member States and private sector companies should develop due diligence and oversight mechanisms to identify, prevent, mitigate and account for how they address the impact of AI systems on the respect for human rights, rule of law and inclusive societies. Member States should also be able to assess the socio-economic impact of AI systems on poverty and ensure that the gap between people living in wealth and poverty, as well as the digital divide among and within countries, are not increased with the massive adoption of AI technologies at present and in the future. In order to do this, in particular, enforceable transparency protocols should be implemented, corresponding to the access to information, including information of public interest held by private entities. Member States, private sector companies and civil society should investigate the sociological and psychological effects of AI-based recommendations on humans in their decision-making autonomy. AI systems identified as potential risks to human rights should be broadly tested by AI actors, including in real-world conditions if needed, as part of the Ethical Impact Assessment, before releasing them in the market.

52. Member States and business enterprises should implement appropriate measures to monitor all phases of an AI system life cycle, including the functioning of algorithms used for decision-making, the data, as well as AI actors involved in the process, especially in public services and where direct end-user interaction is needed, as part of ethical impact assessment. Member States' human rights law obligations should form part of the ethical aspects of AI system assessments.

53. Governments should adopt a regulatory framework that sets out a procedure, particularly for public authorities, to carry out ethical impact assessments on AI systems to predict consequences, mitigate risks, avoid harmful consequences, facilitate citizen participation and address societal challenges. The assessment should also establish appropriate oversight mechanisms, including auditability, traceability and explainability, which enable the assessment of algorithms, data and design processes, as well as include external review of AI systems. Ethical impact assessments should be transparent and open to the public, where appropriate. Such assessments should also be multidisciplinary, multi-stakeholder, multicultural, pluralistic and inclusive. The public authorities should be required to monitor the AI systems implemented and/or deployed by those authorities by introducing appropriate mechanisms and tools.

POLICY AREA 2: ETHICAL GOVERNANCE AND STEWARDSHIP

54. Member States should ensure that AI governance mechanisms are inclusive, transparent, multidisciplinary, multilateral (this includes the possibility of mitigation and redress of harm across borders) and multi-stakeholder. In particular, governance should include aspects of anticipation, and effective protection, monitoring of impact, enforcement and redress.

55. Member States should ensure that harms caused through AI systems are investigated and redressed, by enacting strong enforcement mechanisms and remedial actions, to make certain that human rights and fundamental freedoms and the rule of law are respected in the digital world and in the physical world. Such mechanisms and actions should include remediation mechanisms provided by private and public sector companies. The auditability and traceability of AI systems should be promoted to this end. In addition, Member States should strengthen their institutional capacities to deliver on this commitment and should collaborate with researchers and other stakeholders to investigate, prevent and mitigate any potentially malicious uses of AI systems.

56. Member States are encouraged to develop national and regional AI strategies and to consider forms of soft governance such as a certification mechanism for AI systems and the mutual recognition of their certification, according to the sensitivity of the application domain and expected impact on human rights, the environment and ecosystems, and other ethical considerations set forth in this Recommendation. Such a mechanism might include different levels of audit of systems, data, and adherence to ethical guidelines and to procedural requirements in view of ethical aspects. At the same time, such a mechanism should not hinder innovation or disadvantage small and medium enterprises or start-ups, civil society as well as research and science organizations, as a result of an excessive administrative burden. These mechanisms should also include a regular monitoring component to ensure system robustness and continued integrity and adherence to ethical guidelines over the entire life cycle of the AI system, requiring re-certification if necessary.

57. Member States and public authorities should carry out transparent self-assessment of existing and proposed AI systems, which, in particular, should include the assessment of whether the adoption of AI is appropriate and, if so, should include further assessment to determine what the appropriate method is, as well as assessment as to whether such adoption would result in violations or abuses of Member States' human rights law obligations, and if that is the case, prohibit its use.

58. Member States should encourage public entities, private sector companies and civil society organizations to involve different stakeholders in their AI governance and to consider adding the role of an independent AI Ethics Officer or some other mechanism to oversee ethical impact assessment, auditing and continuous monitoring efforts and ensure ethical guidance of AI systems. Member States, private sector companies and civil society organizations, with the support of UNESCO, are encouraged to create a network of independent AI Ethics Officers to give support to this process at national, regional and international levels.

59. Member States should foster the development of, and access to, a digital ecosystem for ethical and inclusive development of AI systems at the national level, including to address gaps in access to the AI system life cycle, while contributing to international collaboration. Such an ecosystem includes, in particular, digital technologies and infrastructure, and mechanisms for sharing AI knowledge, as appropriate.

60. Member States should establish mechanisms, in collaboration with international organizations, transnational corporations, academic institutions and civil society, to ensure the

active participation of all Member States, especially LMICs, in particular LDCs, LLDCs and SIDS, in international discussions concerning AI governance. This can be through the provision of funds, ensuring equal regional participation, or any other mechanisms. Furthermore, in order to ensure the inclusiveness of AI fora, Member States should facilitate the travel of AI actors in and out of their territory, especially from LMICs, in particular LDCs, LLDCs and SIDS, for the purpose of participating in these fora.

61. Amendments to the existing or elaboration of new national legislation addressing AI systems must comply with Member States' human rights law obligations and promote human rights and fundamental freedoms throughout the AI system life cycle. Promotion thereof should also take the form of governance initiatives, good exemplars of collaborative practices regarding AI systems, and national and international technical and methodological guidelines as AI technologies advance. Diverse sectors, including the private sector, in their practices regarding AI systems must respect, protect and promote human rights and fundamental freedoms using existing and new instruments in combination with this Recommendation.

62. Member States that acquire AI systems for human rights-sensitive use cases, such as law enforcement, welfare, employment, media and information providers, health care and the independent judiciary system should provide mechanisms to monitor the social and economic impact of such systems by appropriate oversight authorities, including independent data protection authorities, sectoral oversight and public bodies responsible for oversight.

63. Member States should enhance the capacity of the judiciary to make decisions related to AI systems as per the rule of law and in line with international law and standards, including in the use of AI systems in their deliberations, while ensuring that the principle of human oversight is upheld. In case AI systems are used by the judiciary, sufficient safeguards are needed to guarantee inter alia the protection of fundamental human rights, the rule of law, judicial independence as well as the principle of human oversight, and to ensure a trustworthy, public interest-oriented and human-centric development and use of AI systems in the judiciary.

64. Member States should ensure that governments and multilateral organizations play a leading role in ensuring the safety and security of AI systems, with multi-stakeholder participation. Specifically, Member States, international organizations and other relevant bodies should develop international standards that describe measurable, testable levels of safety and transparency, so that systems can be objectively assessed and levels of compliance determined. Furthermore, Member States and business enterprises should continuously support strategic research on potential safety and security risks of AI technologies and should encourage research into transparency and explainability, inclusion and literacy by putting additional funding into those areas for different domains and at different levels, such as technical and natural language.

65. Member States should implement policies to ensure that the actions of AI actors are consistent with international human rights law, standards and principles throughout the life cycle of AI systems, while taking into full consideration the current cultural and social diversities, including local customs and religious traditions, with due regard to the precedence and universality of human rights.

66. Member States should put in place mechanisms to require AI actors to disclose and combat any kind of stereotyping in the outcomes of AI systems and data, whether by design or by negligence, and to ensure that training data sets for AI systems do not foster cultural, economic or social inequalities, prejudice, the spreading of disinformation and misinformation, and disruption of freedom of expression and access to information. Particular attention should be given to regions where the data are scarce.

67. Member States should implement policies to promote and increase diversity and inclusiveness that reflect their populations in AI development teams and training datasets, and to ensure equal access to AI technologies and their benefits, particularly for marginalized groups, both from rural and urban zones.

68. Member States should develop, review and adapt, as appropriate, regulatory frameworks to achieve accountability and responsibility for the content and outcomes of AI systems at the different phases of their life cycle. Member States should, where necessary, introduce liability frameworks or clarify the interpretation of existing frameworks to ensure the attribution of accountability for the outcomes and the functioning of AI systems. Furthermore, when developing regulatory frameworks, Member States should, in particular, take into account that ultimate responsibility and accountability must always lie with natural or legal persons and that AI systems should not be given legal personality themselves. To ensure this, such regulatory frameworks should be consistent with the principle of human oversight and establish a comprehensive approach focused on AI actors and the technological processes involved across the different stages of the AI system life cycle.

69. In order to establish norms where these do not exist, or to adapt the existing legal frameworks, Member States should involve all AI actors (including, but not limited to, researchers, representatives of civil society and law enforcement, insurers, investors, manufacturers, engineers, lawyers and users). The norms can mature into best practices, laws and regulations. Member States are further encouraged to use mechanisms such as policy prototypes and regulatory sandboxes to accelerate the development of laws, regulations and policies, including regular reviews thereof, in line with the rapid development of new technologies and ensure that laws and regulations can be tested in a safe environment before being officially adopted. Member States should support local governments in the development of local policies, regulations and laws in line with national and international legal frameworks.

70. Member States should set clear requirements for AI system transparency and explainability so as to help ensure the trustworthiness of the full AI system life cycle. Such requirements should involve the design and implementation of impact mechanisms that take into consideration the nature of application domain, intended use, target audience and feasibility of each particular AI system.

POLICY AREA 3: DATA POLICY

71. Member States should work to develop data governance strategies that ensure the continual evaluation of the quality of training data for AI systems including the adequacy of the data collection and selection processes, proper data security and protection measures, as well as feedback mechanisms to learn from mistakes and share best practices among all AI actors.

72. Member States should put in place appropriate safeguards to protect the right to privacy in accordance with international law, including addressing concerns such as surveillance. Member States should, among others, adopt or enforce legislative frameworks that provide appropriate protection, compliant with international law. Member States should strongly encourage all AI actors, including business enterprises, to follow existing international standards and, in particular, to carry out adequate privacy impact assessments, as part of ethical impact assessments, which take into account the wider socio-economic impact of the intended data processing, and to apply privacy by design in their systems. Privacy should be respected, protected and promoted throughout the life cycle of AI systems.

73. Member States should ensure that individuals retain rights over their personal data and are protected by a framework, which notably foresees: transparency; appropriate safeguards for the processing of sensitive data; an appropriate level of data protection; effective and meaningful accountability schemes and mechanisms; the full enjoyment of the data subjects'

rights and the ability to access and erase their personal data in AI systems, except for certain circumstances in compliance with international law; an appropriate level of protection in full compliance with data protection legislation where data are being used for commercial purposes such as enabling micro-targeted advertising, transferred cross-border; and an effective independent oversight as part of a data governance mechanism which keeps individuals in control of their personal data and fosters the benefits of a free flow of information internationally, including access to data.

74. Member States should establish their data policies or equivalent frameworks, or reinforce existing ones, to ensure full security for personal data and sensitive data, which, if disclosed, may cause exceptional damage, injury or hardship to individuals. Examples include data relating to offences, criminal proceedings and convictions, and related security measures; biometric, genetic and health data; and -personal data such as that relating to race, colour, descent, gender, age, language, religion, political opinion, national origin, ethnic origin, social origin, economic or social condition of birth, or disability and any other characteristics.

75. Member States should promote open data. In this regard, Member States should consider reviewing their policies and regulatory frameworks, including on access to information and open government to reflect AI-specific requirements and promoting mechanisms, such as open repositories for publicly funded or publicly held data and source code and data trusts, to support the safe, fair, legal and ethical sharing of data, among others.

76. Member States should promote and facilitate the use of quality and robust datasets for training, development and use of AI systems, and exercise vigilance in overseeing their collection and use. This could, if possible and feasible, include investing in the creation of gold standard datasets, including open and trustworthy datasets, which are diverse, constructed on a valid legal basis, including consent of data subjects, when required by law. Standards for annotating datasets should be encouraged, including disaggregating data on gender and other bases, so it can easily be determined how a dataset is gathered and what properties it has.

77. Member States, as also suggested in the report of the United Nations Secretary-General's High-level Panel on Digital Cooperation, with the support of the United Nations and UNESCO, should adopt a digital commons approach to data where appropriate, increase interoperability of tools and datasets and interfaces of systems hosting data, and encourage private sector companies to share the data they collect with all stakeholders, as appropriate, for research, innovation or public benefits. They should also promote public and private efforts to create collaborative platforms to share quality data in trusted and secured data spaces.

POLICY AREA 4: DEVELOPMENT AND INTERNATIONAL COOPERATION

78. Member States and transnational corporations should prioritize AI ethics by including discussions of AI-related ethical issues into relevant international, intergovernmental and multi-stakeholder fora.

79. Member States should ensure that the use of AI in areas of development such as education, science, culture, communication and information, health care, agriculture and food supply, environment, natural resource and infrastructure management, economic planning and growth, among others, adheres to the values and principles set forth in this Recommendation.

80. Member States should work through international organizations to provide platforms for international cooperation on AI for development, including by contributing expertise, funding, data, domain knowledge, infrastructure, and facilitating multi-stakeholder collaboration to tackle challenging development problems, especially for LMICs, in particular LDCs, LLDCs and SIDS.

81. Member States should work to promote international collaboration on AI research and innovation, including research and innovation centres and networks that promote greater participation and leadership of researchers from LMICs and other countries, including LDCs, LLDCs and SIDS.

82. Member States should promote AI ethics research by engaging international organizations and research institutions, as well as transnational corporations, that can be a basis for the ethical use of AI systems by public and private entities, including research into the applicability of specific ethical frameworks in specific cultures and contexts, and the possibilities to develop technologically feasible solutions in line with these frameworks.

83. Member States should encourage international cooperation and collaboration in the field of AI to bridge geo-technological lines. Technological exchanges and consultations should take place between Member States and their populations, between the public and private sectors, and between and among the most and least technologically advanced countries in full respect of international law.

POLICY AREA 5: ENVIRONMENT AND ECOSYSTEMS

84. Member States and business enterprises should assess the direct and indirect environmental impact throughout the AI system life cycle, including, but not limited to, its carbon footprint, energy consumption and the environmental impact of raw material extraction for supporting the manufacturing of AI technologies, and reduce the environmental impact of AI systems and data infrastructures. Member States should ensure compliance of all AI actors with environmental law, policies and practices.

85. Member States should introduce incentives, when needed and appropriate, to ensure the development and adoption of rights-based and ethical AI-powered solutions for disaster risk resilience; the monitoring, protection and regeneration of the environment and ecosystems; and the preservation of the planet. These AI systems should involve the participation of local and indigenous communities throughout the life cycle of AI systems and should support circular economy type approaches and sustainable consumption and production patterns. Some examples include using AI systems, when needed and appropriate, to:

(a) Support the protection, monitoring and management of natural resources.

(b) Support the prediction, prevention, control and mitigation of climate-related problems.

(c) Support a more efficient and sustainable food ecosystem.

(d) Support the acceleration of access to and mass adoption of sustainable energy.

(e) Enable and promote the mainstreaming of sustainable infrastructure, sustainable business models and sustainable finance for sustainable development.

(f) Detect pollutants or predict levels of pollution and thus help relevant stakeholders identify, plan and put in place targeted interventions to prevent and reduce pollution and exposure.

86. When choosing AI methods, given the potential data-intensive or resource-intensive character of some of them and the respective impact on the environment, Member States should ensure that AI actors, in line with the principle of proportionality, favour data, energy and resource-efficient AI methods. Requirements should be developed to ensure that appropriate evidence is available to show that an AI application will have the intended effect,

or that safeguards accompanying an AI application can support the justification for its use. If this cannot be done, the precautionary principle must be favoured, and in instances where there are disproportionate negative impacts on the environment, AI should not be used.

POLICY AREA 6: GENDER

87. Member States should ensure that the potential for digital technologies and artificial intelligence to contribute to achieving gender equality is fully maximized, and must ensure that the human rights and fundamental freedoms of girls and women, and their safety and integrity are not violated at any stage of the AI system life cycle. Moreover, Ethical Impact Assessment should include a transversal gender perspective.

88. Member States should have dedicated funds from their public budgets linked to financing gender-responsive schemes, ensure that national digital policies include a gender action plan, and develop relevant policies, for example, on labour education, targeted at supporting girls and women to make sure they are not left out of the digital economy powered by AI. Special investment in providing targeted programmes and gender-specific language, to increase the opportunities of girls' and women's participation in science, technology, engineering, and mathematics (STEM), including information and communication technologies (ICT) disciplines, preparedness, employability, equal career development and professional growth of girls and women, should be considered and implemented.

89. Member States should ensure that the potential of AI systems to advance the achievement of gender equality is realized. They should ensure that these technologies do not exacerbate the already wide gender gaps existing in several fields in the analogue world, and instead eliminate those gaps. These gaps include: the gender wage gap; the unequal representation in certain professions and activities; the lack of representation at top management positions, boards of directors, or research teams in the AI field; the education gap; the digital and AI access, adoption, usage and affordability gap; and the unequal distribution of unpaid work and of the caring responsibilities in our societies.

90. Member States should ensure that gender stereotyping and discriminatory biases are not translated into AI systems, and instead identify and proactively redress these. Efforts are necessary to avoid the compounding negative effect of technological divides in achieving gender equality and avoiding violence such as harassment, bullying or trafficking of girls and women and under-represented groups, including in the online domain.

91. Member States should encourage female entrepreneurship, participation and engagement in all stages of an AI system life cycle by offering and promoting economic, regulatory incentives, among other incentives and support schemes, as well as policies that aim at a balanced gender participation in AI research in academia, gender representation on digital and AI companies' top management positions, boards of directors and research teams. Member States should ensure that public funds (for innovation, research and technologies) are channelled to inclusive programmes and companies, with clear gender representation, and that private funds are similarly encouraged through affirmative action principles. Policies on harassment-free environments should be developed and enforced, together with the encouragement of the transfer of best practices on how to promote diversity throughout the AI system life cycle.

92. Member States should promote gender diversity in AI research in academia and industry by offering incentives to girls and women to enter the field, putting in place mechanisms to fight gender stereotyping and harassment within the AI research community, and encouraging academic and private entities to share best practices on how to enhance gender diversity.

93. UNESCO can help form a repository of best practices for incentivizing the participation of girls, women and under-represented groups in all stages of the AI system life cycle.

POLICY AREA 7: CULTURE

94. Member States are encouraged to incorporate AI systems, where appropriate, in the preservation, enrichment, understanding, promotion, management and accessibility of tangible, documentary and intangible cultural heritage, including endangered languages as well as indigenous languages and knowledges, for example by introducing or updating educational programmes related to the application of AI systems in these areas, where appropriate, and by ensuring a participatory approach, targeted at institutions and the public.

95. Member States are encouraged to examine and address the cultural impact of AI systems, especially natural language processing (NLP) applications such as automated translation and voice assistants, on the nuances of human language and expression. Such assessments should provide input for the design and implementation of strategies that maximize the benefits from these systems by bridging cultural gaps and increasing human understanding, as well as addressing the negative implications such as the reduction of use, which could lead to the disappearance of endangered languages, local dialects, and tonal and cultural variations associated with human language and expression.

96. Member States should promote AI education and digital training for artists and creative professionals to assess the suitability of AI technologies for use in their profession, and contribute to the design and implementation of suitable AI technologies, as AI technologies are being used to create, produce, distribute, broadcast and consume a variety of cultural goods and services, bearing in mind the importance of preserving cultural heritage, diversity and artistic freedom.

97. Member States should promote awareness and evaluation of AI tools among local cultural industries and small and medium enterprises working in the field of culture, to avoid the risk of concentration in the cultural market.

98. Member States should engage technology companies and other stakeholders to promote a diverse supply of and plural access to cultural expressions, and in particular to ensure that algorithmic recommendation enhances the visibility and discoverability of local content.

99. Member States should foster new research at the intersection between AI and intellectual property (IP), for example to determine whether or how to protect with IP rights the works created by means of AI technologies. Member States should also assess how AI technologies are affecting the rights or interests of IP owners, whose works are used to research, develop, train or implement AI applications.

100. Member States should encourage museums, galleries, libraries and archives at the national level to use AI systems to highlight their collections and enhance their libraries, databases and knowledge base, while also providing access to their users.

POLICY AREA 8: EDUCATION AND RESEARCH

101. Member States should work with international organizations, educational institutions and private and non-governmental entities to provide adequate AI literacy education to the public on all levels in all countries in order to empower people and reduce the digital divides and digital access inequalities resulting from the wide adoption of AI systems.

102. Member States should promote the acquisition of “prerequisite skills” for AI education, such as basic literacy, numeracy, coding and digital skills, and media and information literacy, as well as critical and creative thinking, teamwork, communication, socio-emotional and AI ethics skills, especially in countries and in regions or areas within countries where there are notable gaps in the education of these skills.

103. Member States should promote general awareness programmes about AI developments, including on data and the opportunities and challenges brought about by AI technologies, the impact of AI systems on human rights and their implications, including children’s rights. These programmes should be accessible to non-technical as well as technical groups.

104. Member States should encourage research initiatives on the responsible and ethical use of AI technologies in teaching, teacher training and e-learning, among other issues, to enhance opportunities and mitigate the challenges and risks involved in this area. The initiatives should be accompanied by an adequate assessment of the quality of education and impact on students and teachers of the use of AI technologies. Member States should also ensure that AI technologies empower students and teachers and enhance their experience, bearing in mind that relational and social aspects and the value of traditional forms of education are vital in teacher-student and student-student relationships and should be considered when discussing the adoption of AI technologies in education. AI systems used in learning should be subject to strict requirements when it comes to the monitoring, assessment of abilities, or prediction of the learners’ behaviours. AI should support the learning process without reducing cognitive abilities and without extracting sensitive information, in compliance with relevant personal data protection standards. The data handed over to acquire knowledge collected during the learner’s interactions with the AI system must not be subject to misuse, misappropriation or criminal exploitation, including for commercial purposes.

105. Member States should promote the participation and leadership of girls and women, diverse ethnicities and cultures, persons with disabilities, marginalized and vulnerable people or people in vulnerable situations, minorities and all persons not enjoying the full benefits of digital inclusion, in AI education programmes at all levels, as well as the monitoring and sharing of best practices in this regard with other Member States.

106. Member States should develop, in accordance with their national education programmes and traditions, AI ethics curricula for all levels, and promote cross-collaboration between AI technical skills education and humanistic, ethical and social aspects of AI education. Online courses and digital resources of AI ethics education should be developed in local languages, including indigenous languages, and take into account the diversity of environments, especially ensuring accessibility of formats for persons with disabilities.

107. Member States should promote and support AI research, notably AI ethics research, including for example through investing in such research or by creating incentives for the public and private sectors to invest in this area, recognizing that research contributes significantly to the further development and improvement of AI technologies with a view to promoting international law and the values and principles set forth in this Recommendation. Member States should also publicly promote the best practices of, and cooperation with, researchers and companies who develop AI in an ethical manner.

108. Member States should ensure that AI researchers are trained in research ethics and require them to include ethical considerations in their designs, products and publications, especially in the analyses of the datasets they use, how they are annotated, and the quality and scope of the results with possible applications.

109. Member States should encourage private sector companies to facilitate the access of the scientific community to their data for research, especially in LMICs, in particular LDCs, LLDCs and SIDS. This access should conform to relevant privacy and data protection standards.

110. To ensure a critical evaluation of AI research and proper monitoring of potential misuses or adverse effects, Member States should ensure that any future developments with regards to AI technologies should be based on rigorous and independent scientific research, and promote interdisciplinary AI research by including disciplines other than science, technology, engineering and mathematics (STEM), such as cultural studies, education, ethics, international relations, law, linguistics, philosophy, political science, sociology and psychology.

111. Recognizing that AI technologies present great opportunities to help advance scientific knowledge and practice, especially in traditionally model-driven disciplines, Member States should encourage scientific communities to be aware of the benefits, limits and risks of their use; this includes attempting to ensure that conclusions drawn from data-driven approaches, models and treatments are robust and sound. Furthermore, Member States should welcome and support the role of the scientific community in contributing to policy and in cultivating awareness of the strengths and weaknesses of AI technologies.

POLICY AREA 9: COMMUNICATION AND INFORMATION

112. Member States should use AI systems to improve access to information and knowledge. This can include support to researchers, academia, journalists, the general public and developers, to enhance freedom of expression, academic and scientific freedoms, access to information, and increased proactive disclosure of official data and information.

113. Member States should ensure that AI actors respect and promote freedom of expression as well as access to information with regard to automated content generation, moderation and curation. Appropriate frameworks, including regulation, should enable transparency of online communication and information operators and ensure users have access to a diversity of viewpoints, as well as processes for prompt notification to the users on the reasons for removal or other treatment of content, and appeal mechanisms that allow users to seek redress.

114. Member States should invest in and promote digital and media and information literacy skills to strengthen critical thinking and competencies needed to understand the use and implication of AI systems, in order to mitigate and counter disinformation, misinformation and hate speech. A better understanding and evaluation of both the positive and potentially harmful effects of recommender systems should be part of those efforts.

115. Member States should create enabling environments for media to have the rights and resources to effectively report on the benefits and harms of AI systems, and also encourage media to make ethical use of AI systems in their operations.

POLICY AREA 10: ECONOMY AND LABOUR

116. Member States should assess and address the impact of AI systems on labour markets and its implications for education requirements, in all countries and with special emphasis on countries where the economy is labour-intensive. This can include the introduction of a wider range of “core” and interdisciplinary skills at all education levels to provide current workers and new generations a fair chance of finding jobs in a rapidly changing market, and to ensure their awareness of the ethical aspects of AI systems. Skills such as “learning how to learn”, communication, critical thinking, teamwork, empathy, and the ability to transfer one’s knowledge across domains, should be taught alongside specialist, technical skills, as well as

low-skilled tasks. Being transparent about what skills are in demand and updating curricula around these are key.

117. Member States should support collaboration agreements among governments, academic institutions, vocational education and training institutions, industry, workers' organizations and civil society to bridge the gap of skillset requirements to align training programmes and strategies with the implications of the future of work and the needs of industry, including small and medium enterprises. Project-based teaching and learning approaches for AI should be promoted, allowing for partnerships between public institutions, private sector companies, universities and research centres.

118. Member States should work with private sector companies, civil society organizations and other stakeholders, including workers and unions to ensure a fair transition for at-risk employees. This includes putting in place upskilling and reskilling programmes, finding effective mechanisms of retaining employees during those transition periods, and exploring "safety net" programmes for those who cannot be retrained. Member States should develop and implement programmes to research and address the challenges identified that could include upskilling and reskilling, enhanced social protection, proactive industry policies and interventions, tax benefits, new taxation forms, among others. Member States should ensure that there is sufficient public funding to support these programmes. Relevant regulations, such as tax regimes, should be carefully examined and changed if needed to counteract the consequences of unemployment caused by AI-based automation.

119. Member States should encourage and support researchers to analyse the impact of AI systems on the local labour environment in order to anticipate future trends and challenges. These studies should have an interdisciplinary approach and investigate the impact of AI systems on economic, social and geographic sectors, as well as on human-robot interactions and human-human relationships, in order to advise on reskilling and redeployment best practices.

120. Member States should take appropriate steps to ensure competitive markets and consumer protection, considering possible measures and mechanisms at national, regional and international levels, to prevent abuse of dominant market positions, including by monopolies, in relation to AI systems throughout their life cycle, whether these are data, research, technology, or market. Member States should prevent the resulting inequalities, assess relevant markets and promote competitive markets. Due consideration should be given to LMICs, in particular LDCs, LLDCs and SIDS, which are more exposed and vulnerable to the possibility of abuses of market dominance as a result of a lack of infrastructure, human capacity and regulations, among other factors. AI actors developing AI systems in countries which have established or adopted ethical standards on AI should respect these standards when exporting these products, developing or applying their AI systems in countries where such standards may not exist, while respecting applicable international law and domestic legislation, standards and practices of these countries.

POLICY AREA 11: HEALTH AND SOCIAL WELL-BEING

121. Member States should endeavour to employ effective AI systems for improving human health and protecting the right to life, including mitigating disease outbreaks, while building and maintaining international solidarity to tackle global health risks and uncertainties, and ensure that their deployment of AI systems in health care be consistent with international law and their human rights law obligations. Member States should ensure that actors involved in health care AI systems take into consideration the importance of a patient's relationships with their family and with health care staff.

122. Member States should ensure that the development and deployment of AI systems related to health in general and mental health in particular, paying due attention to children and youth, is regulated to the effect that they are safe, effective, efficient, scientifically and medically proven and enable evidence-based innovation and medical progress. Moreover, in the related area of digital health interventions, Member States are strongly encouraged to actively involve patients and their representatives in all relevant steps of the development of the system.

123. Member States should pay particular attention in regulating prediction, detection and treatment solutions for health care in AI applications by:

(a) ensuring oversight to minimize and mitigate bias;

(b) ensuring that the professional, the patient, caregiver or service user is included as a “domain expert” in the team in all relevant steps when developing the algorithms;

(c) paying due attention to privacy because of the potential need for being medically monitored and ensuring that all relevant national and international data protection requirements are met;

(d) ensuring effective mechanisms so that those whose personal data is being analysed are aware of and provide informed consent for the use and analysis of their data, without preventing access to health care;

(e) ensuring the human care and final decision of diagnosis and treatment are taken always by humans while acknowledging that AI systems can also assist in their work;

(f) ensuring, where necessary, the review of AI systems by an ethical research committee prior to clinical use.

124. Member States should establish research on the effects and regulation of potential harms to mental health related to AI systems, such as higher degrees of depression, anxiety, social isolation, developing addiction, trafficking, radicalization and misinformation, among others.

125. Member States should develop guidelines for human-robot interactions and their impact on human-human relationships, based on research and directed at the future development of robots, and with special attention to the mental and physical health of human beings. Particular attention should be given to the use of robots in health care and the care for older persons and persons with disabilities, in education, and robots for use by children, toy robots, chatbots and companion robots for children and adults. Furthermore, assistance of AI technologies should be applied to increase the safety and ergonomic use of robots, including in a human-robot working environment. Special attention should be paid to the possibility of using AI to manipulate and abuse human cognitive biases.

126. Member States should ensure that human-robot interactions comply with the same values and principles that apply to any other AI systems, including human rights and fundamental freedoms, the promotion of diversity, and the protection of vulnerable people or people in vulnerable situations. Ethical questions related to AI-powered systems for neurotechnologies and brain-computer interfaces should be considered in order to preserve human dignity and autonomy.

127. Member States should ensure that users can easily identify whether they are interacting with a living being, or with an AI system imitating human or animal characteristics, and can effectively refuse such interaction and request human intervention.

128. Member States should implement policies to raise awareness about the anthropomorphization of AI technologies and technologies that recognize and mimic human emotions, including in the language used to mention them, and assess the manifestations, ethical implications and possible limitations of such anthropomorphization, in particular in the context of robot-human interaction and especially when children are involved.

129. Member States should encourage and promote collaborative research into the effects of long-term interaction of people with AI systems, paying particular attention to the psychological and cognitive impact that these systems can have on children and young people. This should be done using multiple norms, principles, protocols, disciplinary approaches, and assessment of the modification of behaviours and habits, as well as careful evaluation of the downstream cultural and societal impacts. Furthermore, Member States should encourage research on the effect of AI technologies on health system performance and health outcomes.

130. Member States, as well as all stakeholders, should put in place mechanisms to meaningfully engage children and young people in conversations, debates and decision-making with regard to the impact of AI systems on their lives and futures.

V. MONITORING AND EVALUATION

131. Member States should, according to their specific conditions, governing structures and constitutional provisions, credibly and transparently monitor and evaluate policies, programmes and mechanisms related to ethics of AI, using a combination of quantitative and qualitative approaches. To support Member States, UNESCO can contribute by:

(a) developing a UNESCO methodology for Ethical Impact Assessment (EIA) of AI technologies based on rigorous scientific research and grounded in international human rights law, guidance for its implementation in all stages of the AI system life cycle, and capacity-building materials to support Member States' efforts to train government officials, policy-makers and other relevant AI actors on EIA methodology;

(b) developing a UNESCO readiness assessment methodology to assist Member States in identifying their status at specific moments of their readiness trajectory along a continuum of dimensions;

(c) developing a UNESCO methodology to evaluate *ex ante* and *ex post* the effectiveness and efficiency of the policies for AI ethics and incentives against defined objectives;

(d) strengthening the research- and evidence-based analysis of and reporting on policies regarding AI ethics;

(e) collecting and disseminating progress, innovations, research reports, scientific publications, data and statistics regarding policies for AI ethics, including through existing initiatives, to support sharing best practices and mutual learning, and to advance the implementation of this Recommendation.

132. Processes for monitoring and evaluation should ensure broad participation of all stakeholders, including, but not limited to, vulnerable people or people in vulnerable situations. Social, cultural and gender diversity should be ensured, with a view to improving learning

processes and strengthening the connections between findings, decision-making, transparency and accountability for results.

133. In the interests of promoting best policies and practices related to ethics of AI, appropriate tools and indicators should be developed for assessing the effectiveness and efficiency thereof against agreed standards, priorities and targets, including specific targets for persons belonging to disadvantaged, marginalized populations, and vulnerable people or people in vulnerable situations, as well as the impact of AI systems at individual and societal levels. The monitoring and assessment of the impact of AI systems and related AI ethics policies and practices should be carried out continuously in a systematic way proportionate to the relevant risks. This should be based on internationally agreed frameworks and involve evaluations of private and public institutions, providers and programmes, including self-evaluations, as well as tracer studies and the development of sets of indicators. Data collection and processing should be conducted in accordance with international law, national legislation on data protection and data privacy, and the values and principles outlined in this Recommendation.

134. In particular, Member States may wish to consider possible mechanisms for monitoring and evaluation, such as an ethics commission, AI ethics observatory, repository covering human rights-compliant and ethical development of AI systems, or contributions to existing initiatives by addressing adherence to ethical principles across UNESCO's areas of competence, an experience-sharing mechanism, AI regulatory sandboxes, and an assessment guide for all AI actors to evaluate their adherence to policy recommendations mentioned in this document.

VI. UTILIZATION AND EXPLOITATION OF THE PRESENT RECOMMENDATION

135. Member States and all other stakeholders as identified in this Recommendation should respect, promote and protect the ethical values, principles and standards regarding AI that are identified in this Recommendation, and should take all feasible steps to give effect to its policy recommendations.

136. Member States should strive to extend and complement their own action in respect of this Recommendation, by cooperating with all relevant national and international governmental and non-governmental organizations, as well as transnational corporations and scientific organizations, whose activities fall within the scope and objectives of this Recommendation. The development of a UNESCO Ethical Impact Assessment methodology and the establishment of national commissions for the ethics of AI can be important instruments for this.

VII. PROMOTION OF THE PRESENT RECOMMENDATION

137. UNESCO has the vocation to be the principal United Nations agency to promote and disseminate this Recommendation, and accordingly will work in collaboration with other relevant United Nations entities, while respecting their mandate and avoiding duplication of work.

138. UNESCO, including its bodies, such as the World Commission on the Ethics of Scientific Knowledge and Technology (COMEST), the International Bioethics Committee (IBC) and the Intergovernmental Bioethics Committee (IGBC), will also work in collaboration with other international, regional and sub-regional governmental and non-governmental organizations.

139. Even though, within UNESCO, the mandate to promote and protect falls within the authority of governments and intergovernmental bodies, civil society will be an important actor

to advocate for the public sector's interests and therefore UNESCO needs to ensure and promote its legitimacy.

VIII. FINAL PROVISIONS

140. This Recommendation needs to be understood as a whole, and the foundational values and principles are to be understood as complementary and interrelated.

141. Nothing in this Recommendation may be interpreted as replacing, altering or otherwise prejudicing States' obligations or rights under international law, or as approval for any State, other political, economic or social actor, group or person to engage in any activity or perform any act contrary to human rights, fundamental freedoms, human dignity and concern for the environment and ecosystems, both living and non-living.

Recommendation of the Council on Artificial Intelligence

OECD Legal Instruments

OECD

BETTER POLICIES FOR BETTER LIVES

Please cite this document as:
OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449

Series: OECD Legal Instruments

Photo credit: © kras99/Shutterstock.com

Background Information

The Recommendation on Artificial Intelligence (AI) – the first intergovernmental standard on AI – was adopted by the OECD Council at Ministerial level on 22 May 2019 on the proposal of the Committee on Digital Economy Policy (CDEP). The Recommendation aims to foster innovation and trust in AI by promoting the responsible stewardship of trustworthy AI while ensuring respect for human rights and democratic values. Complementing existing OECD standards in areas such as privacy, digital security risk management, and responsible business conduct, the Recommendation focuses on AI-specific issues and sets a standard that is implementable and sufficiently flexible to stand the test of time in this rapidly evolving field. In June 2019, at the Osaka Summit, G20 Leaders welcomed G20 AI Principles, drawn from the OECD Recommendation.

The Recommendation identifies five complementary values-based principles for the responsible stewardship of trustworthy AI and calls on AI actors to promote and implement them:

- inclusive growth, sustainable development and well-being;
- human-centred values and fairness;
- transparency and explainability;
- robustness, security and safety;
- and accountability.

In addition to and consistent with these value-based principles, the Recommendation also provides five recommendations to policy-makers pertaining to national policies and international co-operation for trustworthy AI, namely:

- investing in AI research and development;
- fostering a digital ecosystem for AI;
- shaping an enabling policy environment for AI;
- building human capacity and preparing for labour market transformation;
- and international co-operation for trustworthy AI.

The Recommendation also includes a provision for the development of metrics to measure AI research, development and deployment, and for building an evidence base to assess progress in its implementation.

The OECD's work on Artificial Intelligence and rationale for developing the OECD Recommendation on Artificial Intelligence

Artificial Intelligence (AI) is a general-purpose technology that has the potential to improve the welfare and well-being of people, to contribute to positive sustainable global economic activity, to increase innovation and productivity, and to help respond to key global challenges. It is deployed in many sectors ranging from production, finance and transport to healthcare and security.

Alongside benefits, AI also raises challenges for our societies and economies, notably regarding economic shifts and inequalities, competition, transitions in the labour market, and implications for democracy and human rights.

The OECD has undertaken empirical and policy activities on AI in support of the policy debate over the past two years, starting with a Technology Foresight Forum on AI in 2016 and an international conference on *AI: Intelligent Machines, Smart Policies* in 2017. The Organisation also conducted analytical and measurement work that provides an overview of the AI technical landscape, maps economic and social impacts of AI technologies and their applications, identifies major policy considerations, and describes AI initiatives from governments and other stakeholders at national and international levels.

This work has demonstrated the need to shape a stable policy environment at the international level to foster trust in and adoption of AI in society. Against this background, the OECD Committee on Digital Economy Policy (CDEP) agreed to develop a draft Council Recommendation to promote a human-centric approach to trustworthy AI, that fosters research, preserves economic incentives to innovate, and applies to all stakeholders.

Complementing existing OECD standards already relevant to AI – such as those on privacy and data protection, digital security risk management, and responsible business conduct – the Recommendation focuses on policy issues that are specific to AI and strives to set a standard that is implementable and flexible enough to stand the test of time in a rapidly evolving field. The Recommendation contains five high-level values-based principles and five recommendations for national policies and international co-operation. It also proposes a common understanding of key terms, such as "AI system" and "AI actors", for the purposes of the Recommendation.

More specifically, the Recommendation includes two substantive sections:

1. **Principles for responsible stewardship of trustworthy AI**: the first section sets out five complementary principles relevant to all stakeholders: *i)* inclusive growth, sustainable development and well-being; *ii)* human-centred values and fairness; *iii)* transparency and explainability; *iv)* robustness, security and safety; and *v)* accountability. This section further calls on AI actors to promote and implement these principles according to their roles.
2. **National policies and international co-operation for trustworthy AI**: consistent with the five aforementioned principles, this section provides five recommendations to Members and non-Members having adhered to the draft Recommendation (hereafter the "Adherents") to implement in their national policies and international co-operation: *i)* investing in AI research and development; *ii)* fostering a digital ecosystem for AI; *iii)* shaping an enabling policy environment for AI; *iv)* building human capacity and preparing for labour market transformation; and *v)* international co-operation for trustworthy AI.

An inclusive and participatory process for developing the Recommendation

The development of the Recommendation was participatory in nature, incorporating input from a broad range of sources throughout the process. In May 2018, the CDEP agreed to form an expert group to scope principles to foster trust in and adoption of AI, with a view to developing a draft Council Recommendation in the course of 2019. The AI Group of experts at the OECD (AIGO) was subsequently established, comprising over 50 experts from different disciplines and different sectors (government, industry, civil society, trade unions, the technical community and academia) - see http://www.oecd.org/going-digital/ai/oecd-aigo-membership-list.pdf for the full list. Between September 2018 and February 2019 the group held four meetings: in Paris, France, in September and November 2018, in Cambridge, MA, United States, at the Massachusetts Institute of Technology (MIT) in January 2019, back to back with the MIT AI Policy Congress, and finally in Dubai, United Arab Emirates, at the World Government Summit in February 2019. The work benefited from the diligence, engagement and substantive contributions of the experts participating in AIGO, as well as from their multi-stakeholder and multidisciplinary backgrounds.

Drawing on the final output document of the AIGO, a draft Recommendation was developed in the CDEP and with the consultation of other relevant OECD bodies. The CDEP approved a final draft Recommendation and agreed to transmit it to the OECD Council for adoption in a special meeting on 14-15 March 2019. The OECD Council adopted the Recommendation at its meeting at Ministerial level on 22-23 May 2019.

Follow-up, monitoring of implementation and dissemination tools

The OECD Recommendation on AI provides the first intergovernmental standard for AI policies and a foundation on which to conduct further analysis and develop tools to support governments in their implementation efforts. In this regard, it instructs the CDEP to monitor the implementation of the Recommendation and report to the Council on its implementation and continued relevance five years after its adoption and regularly thereafter. The CDEP is also instructed to continue its work on AI, building on this Recommendation, and taking into account work in other international fora, such as UNESCO, the European Union, the Council of Europe and the initiative to build an International Panel on AI (see https://pm.gc.ca/eng/news/2018/12/06/mandate-international-panel-artificial-intelligence and https://www.gouvernement.fr/en/france-and-canada-create-new-expert-international-panel-on-artificial-intelligence).

In order to support implementation of the Recommendation, the Council instructed the CDEP to develop practical guidance for implementation, to provide a forum for exchanging information on AI policy and activities, and to foster multi-stakeholder and interdisciplinary dialogue. This will be achieved largely through the OECD AI Policy Observatory, an inclusive hub for public policy on AI that aims to help countries encourage, nurture and monitor the responsible development of trustworthy artificial intelligence systems for the benefit of society. It will combine resources from across the OECD with those of partners from all stakeholder groups to provide multidisciplinary, evidence-based policy analysis on AI. The Observatory is planned to be launched late 2019 and will include a live database of AI strategies, policies and initiatives that countries and other stakeholders can share and update, enabling the comparison of their key elements in an interactive manner. It will also be continuously updated with AI metrics, measurements, policies and good practices that could lead to further updates in the practical guidance for implementation.

The Recommendation is open to non-OECD Member adherence, underscoring the global relevance of OECD AI policy work as well as the Recommendation's call for international co-operation.

Artificial Intelligence (AI) tools and systems can support countries in their response to the COVID-19 crisis. For example, AI can help policymakers and the medical community understand the COVID-19 virus and accelerate research on treatments by rapidly analysing large volumes of research data. It can also be employed to help detect, diagnose and prevent the spread of the virus. Conversational and interactive AI systems help respond to the health crisis through personalised information, advice and treatment. Finally, AI tools can help monitor the economic crisis and the recovery – for example, via satellite, social networking and other data (e.g. Google's Community Mobility Reports) – and can help learn from the crisis and build early warning system for future outbreaks. However, in order to make the most of these innovative solutions, AI systems need to be designed, developed and deployed in a trustworthy manner, consistent with the Recommendation: they should respect human rights and privacy; be transparent, explainable, robust, secure and safe; and actors involved in their development and use should remain accountable.

For more information, see:

- Using artificial intelligence to help combat COVID-19;
- Tracking and tracing COVID: Protecting privacy and data while using apps and biometrics

For further information please consult: oecd.ai.

Contact information: ai@oecd.org.

THE COUNCIL,

HAVING REGARD to Article 5 b) of the Convention on the Organisation for Economic Co-operation and Development of 14 December 1960;

HAVING REGARD to the OECD Guidelines for Multinational Enterprises [OECD/LEGAL/0144]; Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data [OECD/LEGAL/0188]; Recommendation of the Council concerning Guidelines for Cryptography Policy [OECD/LEGAL/0289]; Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information [OECD/LEGAL/0362]; Recommendation of the Council on Digital Security Risk Management for Economic and Social Prosperity [OECD/LEGAL/0415]; Recommendation of the Council on Consumer Protection in E-commerce [OECD/LEGAL/0422]; Declaration on the Digital Economy: Innovation, Growth and Social Prosperity (Cancún Declaration) [OECD/LEGAL/0426]; Declaration on Strengthening SMEs and Entrepreneurship for Productivity and Inclusive Growth [OECD/LEGAL/0439]; as well as the 2016 Ministerial Statement on Building more Resilient and Inclusive Labour Markets, adopted at the OECD Labour and Employment Ministerial Meeting;

HAVING REGARD to the Sustainable Development Goals set out in the 2030 Agenda for Sustainable Development adopted by the United Nations General Assembly (A/RES/70/1) as well as the 1948 Universal Declaration of Human Rights;

HAVING REGARD to the important work being carried out on artificial intelligence (hereafter, “AI”) in other international governmental and non-governmental fora;

RECOGNISING that AI has pervasive, far-reaching and global implications that are transforming societies, economic sectors and the world of work, and are likely to increasingly do so in the future;

RECOGNISING that AI has the potential to improve the welfare and well-being of people, to contribute to positive sustainable global economic activity, to increase innovation and productivity, and to help respond to key global challenges;

RECOGNISING that, at the same time, these transformations may have disparate effects within, and between societies and economies, notably regarding economic shifts, competition, transitions in the labour market, inequalities, and implications for democracy and human rights, privacy and data protection, and digital security;

RECOGNISING that trust is a key enabler of digital transformation; that, although the nature of future AI applications and their implications may be hard to foresee, the trustworthiness of AI systems is a key factor for the diffusion and adoption of AI; and that a well-informed whole-of-society public debate is necessary for capturing the beneficial potential of the technology, while limiting the risks associated with it;

UNDERLINING that certain existing national and international legal, regulatory and policy frameworks already have relevance to AI, including those related to human rights, consumer and personal data protection, intellectual property rights, responsible business conduct, and competition, while noting that the appropriateness of some frameworks may need to be assessed and new approaches developed;

RECOGNISING that given the rapid development and implementation of AI, there is a need for a stable policy environment that promotes a human-centric approach to trustworthy AI, that fosters research, preserves economic incentives to innovate, and that applies to all stakeholders according to their role and the context;

CONSIDERING that embracing the opportunities offered, and addressing the challenges raised, by AI applications, and empowering stakeholders to engage is essential to fostering adoption of trustworthy AI in society, and to turning AI trustworthiness into a competitive parameter in the global marketplace;

On the proposal of the Committee on Digital Economy Policy:

I. AGREES that for the purpose of this Recommendation the following terms should be understood as follows:

- *AI system*: An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.
- *AI system lifecycle*: AI system lifecycle phases involve: *i)* 'design, data and models'; which is a context-dependent sequence encompassing planning and design, data collection and processing, as well as model building; *ii)* 'verification and validation'; *iii)* 'deployment'; and *iv)* 'operation and monitoring'. These phases often take place in an iterative manner and are not necessarily sequential. The decision to retire an AI system from operation may occur at any point during the operation and monitoring phase.
- *AI knowledge*: AI knowledge refers to the skills and resources, such as data, code, algorithms, models, research, know-how, training programmes, governance, processes and best practices, required to understand and participate in the AI system lifecycle.
- *AI actors*: AI actors are those who play an active role in the AI system lifecycle, including organisations and individuals that deploy or operate AI.
- *Stakeholders*: Stakeholders encompass all organisations and individuals involved in, or affected by, AI systems, directly or indirectly. AI actors are a subset of stakeholders.

Section 1: Principles for responsible stewardship of trustworthy AI

II. RECOMMENDS that Members and non-Members adhering to this Recommendation (hereafter the "Adherents") promote and implement the following principles for responsible stewardship of trustworthy AI, which are relevant to all stakeholders.

III. CALLS ON all AI actors to promote and implement, according to their respective roles, the following Principles for responsible stewardship of trustworthy AI.

IV. UNDERLINES that the following principles are complementary and should be considered as a whole.

1.1. Inclusive growth, sustainable development and well-being

Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as augmenting human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being.

1.2. Human-centred values and fairness

a) AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights.

b) To this end, AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of art.

1.3. Transparency and explainability

AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art:

i. to foster a general understanding of AI systems,

ii. to make stakeholders aware of their interactions with AI systems, including in the workplace,

iii. to enable those affected by an AI system to understand the outcome, and,

iv. to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.

1.4. Robustness, security and safety

a) AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk.

b) To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art.

c) AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.

1.5. Accountability

AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art.

Section 2: National policies and international co-operation for trustworthy AI

V. RECOMMENDS that Adherents implement the following recommendations, consistent with the principles in section 1, in their national policies and international co-operation, with special attention to small and medium-sized enterprises (SMEs).

2.1. Investing in AI research and development

a) Governments should consider long-term public investment, and encourage private investment, in research and development, including interdisciplinary efforts, to spur innovation in trustworthy AI that focus on challenging technical issues and on AI-related social, legal and ethical implications and policy issues.

b) Governments should also consider public investment and encourage private investment in open datasets that are representative and respect privacy and data protection to support an environment for AI research and development that is free of inappropriate bias and to improve interoperability and use of standards.

2.2. Fostering a digital ecosystem for AI

Governments should foster the development of, and access to, a digital ecosystem for trustworthy AI. Such an ecosystem includes in particular digital technologies and infrastructure, and mechanisms for sharing AI

knowledge, as appropriate. In this regard, governments should consider promoting mechanisms, such as data trusts, to support the safe, fair, legal and ethical sharing of data.

2.3. Shaping an enabling policy environment for AI

a) Governments should promote a policy environment that supports an agile transition from the research and development stage to the deployment and operation stage for trustworthy AI systems. To this effect, they should consider using experimentation to provide a controlled environment in which AI systems can be tested, and scaled-up, as appropriate.

b) Governments should review and adapt, as appropriate, their policy and regulatory frameworks and assessment mechanisms as they apply to AI systems to encourage innovation and competition for trustworthy AI.

2.4. Building human capacity and preparing for labour market transformation

a) Governments should work closely with stakeholders to prepare for the transformation of the world of work and of society. They should empower people to effectively use and interact with AI systems across the breadth of applications, including by equipping them with the necessary skills.

b) Governments should take steps, including through social dialogue, to ensure a fair transition for workers as AI is deployed, such as through training programmes along the working life, support for those affected by displacement, and access to new opportunities in the labour market.

c) Governments should also work closely with stakeholders to promote the responsible use of AI at work, to enhance the safety of workers and the quality of jobs, to foster entrepreneurship and productivity, and aim to ensure that the benefits from AI are broadly and fairly shared.

2.5. International co-operation for trustworthy AI

a) Governments, including developing countries and with stakeholders, should actively co-operate to advance these principles and to progress on responsible stewardship of trustworthy AI.

b) Governments should work together in the OECD and other global and regional fora to foster the sharing of AI knowledge, as appropriate. They should encourage international, cross-sectoral and open multi-stakeholder initiatives to garner long-term expertise on AI.

c) Governments should promote the development of multi-stakeholder, consensus-driven global technical standards for interoperable and trustworthy AI.

d) Governments should also encourage the development, and their own use, of internationally comparable metrics to measure AI research, development and deployment, and gather the evidence base to assess progress in the implementation of these principles.

VI. **INVITES** the Secretary-General and Adherents to disseminate this Recommendation.

VII. **INVITES** non-Adherents to take due account of, and adhere to, this Recommendation.

VIII. **INSTRUCTS** the Committee on Digital Economy Policy:

a) to continue its important work on artificial intelligence building on this Recommendation and taking into account work in other international fora, and to further develop the measurement framework for evidence-based AI policies;

b) to develop and iterate further practical guidance on the implementation of this Recommendation, and to report to the Council on progress made no later than end December 2019;

c) to provide a forum for exchanging information on AI policy and activities including experience with the implementation of this Recommendation, and to foster multi-stakeholder and interdisciplinary dialogue to promote trust in and adoption of AI; and

d) to monitor, in consultation with other relevant Committees, the implementation of this Recommendation and report thereon to the Council no later than five years following its adoption and regularly thereafter.

About the OECD

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD Member countries are: Australia, Austria, Belgium, Canada, Chile, Colombia, Costa Rica, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Türkiye, the United Kingdom and the United States. The European Union takes part in the work of the OECD.

OECD Legal Instruments

Since the creation of the OECD in 1961, around 460 substantive legal instruments have been developed within its framework. These include OECD Acts (i.e. the Decisions and Recommendations adopted by the OECD Council in accordance with the OECD Convention) and other legal instruments developed within the OECD framework (e.g. Declarations, international agreements).

All substantive OECD legal instruments, whether in force or abrogated, are listed in the online Compendium of OECD Legal Instruments. They are presented in five categories:

- **Decisions** are adopted by Council and are legally binding on all Members except those which abstain at the time of adoption. They set out specific rights and obligations and may contain monitoring mechanisms.

- **Recommendations** are adopted by Council and are not legally binding. They represent a political commitment to the principles they contain and entail an expectation that Adherents will do their best to implement them.

- **Substantive Outcome Documents** are adopted by the individual listed Adherents rather than by an OECD body, as the outcome of a ministerial, high-level or other meeting within the framework of the Organisation. They usually set general principles or long-term goals and have a solemn character.

- **International Agreements** are negotiated and concluded within the framework of the Organisation. They are legally binding on the Parties.

- **Arrangement, Understanding and Others**: several other types of substantive legal instruments have been developed within the OECD framework over time, such as the Arrangement on Officially Supported Export Credits, the International Understanding on Maritime Transport Principles and the Development Assistance Committee (DAC) Recommendations.