The European Legislation on AI: a Brief Analysis of its Philosophical Approach

Luciano Floridi[1,2]

Some European legislation on artificial intelligence (AI) had been expected at least since 16 July 2019. On that date, Ursula von der Leyen had pledged that, within 100 days of her election as President of the European Commission, she would have proposed new legislation on AI.[1] At that time, I remarked that it was a reasonable strategy but an unrealistic timeline. The High-Level Expert Group on AI (HLEG, of which I was a member),[2] organised by the European Commission, had only recently published its *Ethics Guidelines for Trustworthy AI* (HLEGAI, 2019) and its *Policy and Investment Recommendations for Trustworthy AI* (HLEGAI, 2019). It seemed evident that the next step would have been the translation of those guidelines and recommendations into a legal framework (Floridi, 2019a). However, the work carried out by the HLEG had also shown that the road ahead was going to be long and laborious. I figured it would have taken at least a year, not three months. I was optimistic. On 19 February 2020, the Commission published the *White Paper on AI—A European Approach to Excellence and Trust* (European Commission, 2020). The document outlined a risk-based approach to AI and policies to promote the uptake of such technology. But, meanwhile, the COVID-19 pandemic had begun to spread, with its deadly effects and immense disruptions.[3] Despite this, on 21 April, 2021, the European Commission published the proposal of the new EU *Artificial Intelligence Act* (henceforth AIA), or, to use its full name, the *Proposal for a regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts* (Artificial Intelligence Act 21, 2021).

According to the European Data Protection Supervisor website, the AIA is "the first initiative, worldwide, that provides a legal framework for Artificial Intelligence

[1] https://ec.europa.eu/commission/presscorner/detail/en/ip_20_403

[2] https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

[3] https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020

✉ Luciano Floridi
luciano.floridi@oii.ox.ac.uk

[1] Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3JS, UK

[2] The Alan Turing Institute, 96 Euston Road, London NW1 2DB, UK

(AI)".[4] Regardless of whether this may be true (see, for example, the US National AI Initiative Act, which became law on 1 January 2021), the AIA is one of the most influential regulatory steps taken so far internationally. On the whole, it is a good starting point to ensure that the development of AI in the EU is ethically sound, legally acceptable, socially equitable, and environmentally sustainable, with a vision of AI that seeks to support the economy, society, and the environment. This is no small ambition, and it will take time and effort to reach a final text that can come close to fulfil it. Yet, the ambition, like von der Leyen's pledge, remains substantially reasonable because the EU is ideally placed to deliver such a normative framework.

Of course, the technical and legal aspects of the AIA will evolve as the proposal goes to the European Parliament and the Council of the European Union, that is, the EU legislator, for further consideration and debate. This is a proposal, and it may take a couple of years before it will be finalised and become binding (the process took 4 years for the GDPR (General Data Protection Regulation 27, 2016), followed by an implementation period of 2 years). However, the underlying philosophy is commendable, despite some limitations. In what follows, I shall highlight some foundational aspects that seem to be more significant. Just a final note of clarification: there are already plenty of short summaries of the proposed legislation, and I shall not provide another one here. Nor shall I comment on some obvious shortcomings already identified by many, from the definition of AI (always a problem for anyone approaching the subject, (Floridi, forthcoming #32)), to the more or less complete and loop-holy list of AI uses or technologies that should be banned. What I am interested in analysing is the philosophy behind the proposal.

From a general perspective, the AIA is a "regulation", not a "directive", so, like the GDPR, it will enter into force on a set date in all 27 Member States, and it will have binding legal force throughout the EU (a directive only indicates goals that each Member State must achieve but does not indicate how to transpose the goals into national laws). Moreover, AIA provisions conferring rights upon individuals and having a sufficiently clear, precise, and unconditional content will enjoy a direct effect, meaning that natural or legal persons will have the possibility to invoke those rights before national courts and tribunals. Also, like the GDPR, the AIA is "extraterritorial", in the more technical sense that it has *territorial extension* (Scott, 2014), or, as I would prefer to put it, it is *aterritorial*, a concept much more coherent with the other three concepts (see below) of *Brussels effect*, *digital constitutionalism*, and *digital sovereignty*. It assumes a post-Westphalian world in which the territoriality of the law no longer applies automatically and may be irrelevant (Floridi, 2014). What counts is whether an AI system or service—for example, a loan management program based on machine learning—has an impact on European citizens, not where the company that provides it or uses it is located, whether physically or legally. This unified, post-Westphalian approach is likely to have several positive effects. The EU presents itself as a single interlocutor, not only in the management of personal data

[4] The quote comes from https://edps.europa.eu/press-publications/press-news/press-releases/2021/artificial-intelligence-act-welcomed-initiative_en

(GDPR), but now also for AI applications. AI companies and vendors[5] will have to deal with the EU, not with individual Member States, when they will have to prove that they comply with the new legislation. Each Member States will appoint a national authority responsible for supervising AI. However, a new European Artificial Intelligence Board (EAIB, akin to the GDPR's European Data Protection Board[6])—consisting of representatives from every Member State, the European Data Protection Supervisor and the Commission—will assist national supervisory authorities and EU lawmakers to ensure the consistent application of the AI Regulation, e.g. regarding the list of prohibited AI practices and high-risk systems. Companies offering AI services are unlikely to be able to ignore a market of 450 million people that accounts for about 1/6 of the world economy, so they will comply (there are steep administrative fines for different kinds of infringements, from up to €10 M or 2% of worldwide annual turnover whichever is higher, to €30 M or 6% of annual worldwide turnover, whichever is higher). This will further extend the so-called *Brussel effect* (Bradford, 2020), whereby companies end up complying with EU regulations even in other countries because it is more practical to have a single approach globally, enabling the EU to extend de facto (though not *de jure*) its laws internationally, through market mechanisms. Companies will also find it more difficult to explain why they do not adopt standards just as high when operating in other countries (Floridi, 2019c). Furthermore, the AIA will place the EU in a position of "leadership by example" for the good governance of AI technologies, especially when interacting with other countries at the forefront in the field of AI research and development (R&D), like Canada,[7] Israel, Japan, Singapore, South Korea, and the UK. In this case too, one may expect a "harmonising" effect similar to that caused by the GDPR. Collaboration between the EU and other countries will be easier thanks to an explicit and reliable legislative framework. As for China and the USA, they will not adopt the same EU approach, but they will need to find ways of collaborating with the EU. China may take inspiration from the AIA to develop its legislation tailored to its approach to AI (Roberts et al., 2021, Roberts et al., 2021). The USA is more likely to adopt an antitrust approach[8] (Cath et al., 2018; Roberts et al., forthcoming), but the AIA may influence state-level legislation, as it already happened with the California Consumer Privacy Act (Barrett, 2019). A future EU-US Trade and Technology Council[9] may also provide a shared platform. More generally, AI could become a subject of negotiation within the framework of the contractual relations that the EU is establishing with third countries.

[5] https://www.datamation.com/artificial-intelligence/ai-companies/

[6] The Commission may have more influnce on the EAIB than it has on the EDPB.

[7] A Regulatory Framework for AI: Recommendations for PIPEDA Reform https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/completed-consultations/consultation-ai/reg-fw_202011/

[8] Aiming for truth, fairness, and equity in your company's use of AI: https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai

[9] Joint communication to the European Parliament, the European Council, and the Council – a new EU-US agenda for global change: https://ec.europa.eu/info/sites/default/files/joint-communication-eu-us-agenda_en.pdf

From an ethical perspective, the AIA inherits the same foundational approach seen in the GDPR: it is based on protecting human dignity and fundamental rights. This is a very positive feature, even if the current proposal of the AIA seems to be somewhat more top-down, less flexible, and less focused on the protection of citizens and their rights than the GDPR. Unfortunately, the AIA uses an anachronistic terminology to define this approach as "human-centric", that is, as an approach that places humanity at the centre of technological development. Yet this is both trivially true and dangerously ambiguous. On the one hand, it is obvious that any technology, AI included, must be at the service of humanity, its values, and needs. On the other hand, one must also consider the environment as crucially important, yet "human-centric" seems to be synonymous with "anthropocentric", and we know how much the planet has suffered from humanity's obsession with its importance and centrality, as if everything must always be at its service, including every aspect of the natural world, no matter at what costs and losses. Fortunately, despite the unfortunate and obsolete terminology, the underlying vision is sound: the AIA emphasises the value of AI as a technology that can be very "green" and provide extraordinary support against pollution and climate change and for the sustainable development of information societies (Cowls et al., 2021). In this, the AIA's approach strengthens the idea that protecting the environment must be a cross-cutting issue at the EU level. It is what I have called the *Green and the Blue human project* (Floridi, 2019b; Floridi & Nobre, 2020), which the EU can and should promote in the world.

Still in terms of ethical approach, the AIA explicitly adopts the ethical guidelines proposed by the HLEG and seeks to eliminate or mitigate the risks of AI, support public trust in these innovative technologies, and further the development and adoption of AI in the EU. This risk-based approach seems convincing (it is a common approach for internal market-based legislation) and aligned with the view that ethics benefits the market, not vice versa. But precisely for this reason, one may argue that the AIA could do much more to protect consumers' rights and be much more incisive about providing measures to redress the possible harms or losses that AI systems may cause. This is the part where one may expect and welcome more improvements in the proposal. It was one of the main recommendations made by the AI4People project: "7. Develop a redress process or mechanism to remedy or compensate for a wrong" (Floridi et al., 2018).

From a technological perspective, one must also praise what is *not* in the proposal. Following an already established approach,[10] the AIA avoids any sci-fi speculations about AI. Unfortunately, scaremongering stories have irresponsibly distracted not only the public (Floridi, 2016), but initially also the work of the European Parliament (European Parliament, 16 February 2017), which experts had criticised in an open letter to the European Commission,[11] and even of the HLEG, where I was among those who strongly insisted about removing non-scientific statements in the first draft to fanciful things such as "artificial consciousness" or AI with "subjective

[10] https://www.euractiv.com/section/digital/opinion/the-eu-is-right-to-refuse-legal-personality-for-artificial-intelligence/

[11] https://www.politico.eu/wp-content/uploads/2018/04/RoboticsOpenLetter.pdf

experience" (HLEGAI 18 December, 2018). Correctly, the AIA treats AI as a technology for solving problems and performing tasks, not as some kind of Frankenstein's monster. Therefore, the proposal excludes the possibility of assigning to AI systems any status as a legal person, with rights and duties, such as the possibility of owning property, entering into contracts, suing and being sued, and so forth (Floridi & Taddeo, 2018). The responsibility of any AI system rests entirely with the people who design, manufacture, market, and use it. Coherently, the proposal stresses the importance of human oversight throughout the text.

I mentioned above some limits. This is a legislative proposal, and it is too early to indicate how it will be revised, but some conceptual improvements (I am not talking about legal and technical issues) may be in order.

At times, the text is ambiguous. For example, the definition of high-risk AI systems, a vital concept of the AIA, could be improved. Some AI systems are discussed as low-risk or zero-risk, and as such, they seem to fall outside the scope of strict compliance and be subject only to voluntary codes of conduct, yet it is unclear how this taxonomy will work in practice, leaving too much room for uncertainty and loopholes. And in the case of high risk, the proposal explicitly combines two senses that it would be preferable to distinguish. On the one hand, there are AI systems that are high-risk because vital issues depend on their proper functioning. Think of an autonomous driving system: it is a "good thing" that should not fail to work. On the other hand, there are AI systems that are high-risk because, if they are used unethically, they can cause significant troubles; think of the abuse of remote, real-time biometric identification for law enforcement purposes, a kind of technological surveillance banned by the proposal. This is a "bad thing" that should not be put into operation. If one does not distinguish between these two senses of high-risk system—something is high risk if it fails to work vs. something is high risk if it is put to work—then one may end confusing the *resilience* that "good" AI systems must have, with the *resistance* that must be exerted towards the "bad" AI systems. Note that conceptual confusion and uncertainty about the specific nature of the risks involved in the design, development, and deployment of AI systems will also affect the feasibility of any conformity assessment (auditing, see Floridi et al. (2018), Mokander and Floridi (2021), Mökander et al. (forthcoming)), a crucial element in the AIA and the certification system it proposes.

In other cases, the proposal is vague, such as when it comes to banning the use of AI systems intended to distort human behaviour, with probable physical or psychological harm. The intent is commendable, but it might risk banning even unproblematic AI systems if this approach were applied in a Draconian way.

Finally, some expectations in the proposal seem too idealistic. For example, consider the properties that the databases used for training machine learning models should satisfy. "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" (Recital 44, note that Recitals have only an interpretative function). These are characteristics highly desirable but rarely met in full. Think, for instance, about the incompleteness and incorrectness of any public database. Therefore, it seems preferable to speak of thresholds below which failure to satisfy these characteristics would be unacceptable (note that I am assuming that "sufficiently" in the

Recital modifies only "relevant" and not all the following properties—as in "sufficiently representative" and so forth—because if this is not the case, then the whole text is too vague).

The AIA will be added to the GDPR and, over time, to the Digital Services Act (Digital Services Act, 15 December 2020) and the Digital Markets Act (Digital Markets Act, 15 December 2020), which, once adopted, will regulate online platforms and services. When this "legislative square" will be complete—make it a pentagon if you add the Data Governance Act (Data Governance Act, 25 November 2020), or a hexagram, with the announced European Health Data Space legislative proposal[12]—the EU will have developed a "digital constitutionalism" (Celeste, 2019; De Gregorio, 2021) for an infosphere where its citizens may live and work better and more sustainably. As I have been arguing for a while, it is clear that the challenge is no longer digital innovation but the governance of the digital, and hence the new morphology of power (Floridi, 2015) and the shaping of digital sovereignty (Floridi, 2020). In tackling these normative challenges, the EU is not simply ahead; it has no competition. There remains an indirect risk to be stressed. The new legislation may not improve but merely push out of the EU some risky AI R&D and its related ethical-legal problems, inviting companies to develop their products and services in other countries where legislation is absent, or less stringent, or not enforced, whilst the EU turns a wilful blind eye—or just inadequately enforces its legislation—and imports services or products obtained elsewhere, checking their current compliance but not their problematic origin (think of an AI-based medical device originally trained on personal data without respecting the GDPR or the AIA). Not respecting this *atemporality* (it should not matter *when* unethical steps were taken) would contradict the *aterritoriality* of the legislation. In this case, the recommendation is obvious: the EU must keep both eyes open and apply its ethical and legal requirements consistently and without hypocrisy, not just to the *status quo*, but also to the *history* of what comes from other places (note that the proposal already goes in this direction). After all, the EU founding Treaties state that, in its relations with the broader world, the EU should uphold and promote its values, contributing to the protection of human rights (Art. 3(5) TEU). It may be technically challenging, expensive, and even internationally problematic, but nobody ever said that doing the right thing was going to be cheap and easy.

Acknowledgements I wish to thank Federico Casolari, Emmie Hine, Joshua Jaffe, Francesca Mazzi, Oreste Pollicino, Huw Roberts, and Paul Timmers for their very helpful feedback on a previous version of this article. The article is really much better thanks to them and any remaining shortcomings are only mine.

[12] https://ec.europa.eu/health/ehealth/dataspace_en

References

Artificial Intelligence Act. (21 April 2021). "Proposal for a regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts." *EUR-Lex - 52021PC0206* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELLAR:e0649735-a372-11eb-9585-01aa75ed71a1.

Barrett, C. (2019). Are the EU GDPR and the California CCPA becoming the de facto global standards for data privacy and protection? *Scitech Lawyer, 15*(3), 24–29.

Bradford, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press.

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial Intelligence and the 'Good Society': The US, EU, and UK approach. *Science and Engineering Ethics, 24*(2), 505–528.

Celeste, E. (2019). Digital constitutionalism: A new systematic theorisation' (2019). *International Review of Law, Computers & Technology, 33*, 76–99.

Cowls, J., Tsamados, A., Taddeo, M., Floridi, L. (2021). "The AI Gambit—Leveraging artificial intelligence to combat climate change: Opportunities, challenges, and recommendations." *SSRN*.

Data Governance Act. (25 November 2020). "Proposal for a regulation of the European Parliament and the Council on European data governance (Data Governance Act)." *EUR-Lex - 52020PC0767* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767.

De Gregorio, Giovanni. (2021). The rise of digital constitutionalism in the European Union. *International Journal of Constitutional Law*. https://doi.org/10.1093/icon/moab001

Digital Markets Act. (15 December 2020). Proposal for a regulation of the European Parliament and the Council on contestable and fair markets in the digital sector (Digital Markets Act). *EUR-Lex - 52020PC0842* https://eur-lex.europa.eu/legal-content/en/ALL/?uri=COM:2020:842:FIN.

Digital Services Act. (15 December 2020). Proposal for a regulation of the European Parliament and the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. *EUR-Lex - 52020PC0825* https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52020PC0825.

European Commission. (19 February 2020). White Paper on AI - A European approach to excellence and trust. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

European Parliament. (16 February 2017). European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html.

Floridi, L. (2014). *The Fourth Revolution - How the infosphere is reshaping human reality*. Oxford University Press.

Floridi, L. (2015). The new grey power. *Philosophy & Technology, 28*(3), 329–332.

Floridi, L. (2016). "Should we be afraid of AI?" *Aeon Essays* https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible.

Floridi, L. (2019a). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence, 1*(6), 261–262.

Floridi, L. (2019b). "The green and the blue: naïve ideas to improve politics in a mature information society." In *The 2018 yearbook of the digital ethics lab*, 183–221. Springer.

Floridi, L. (2019c). Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy & Technology, 32*(2), 185–193.

Floridi, L. (2020). The fight for digital sovereignty: What it is, and why it matters, especially for the EU. *Philosophy & Technology, 33*(3), 369–378.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI Society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707.

Floridi, Luciano, and Kia Nobre. 2020. The green and the blue: How AI may be a force for good. OECD - available online.

Floridi, L., & Taddeo, M. (2018). Romans would have denied robots legal personhood. *Nature, 557*(7705), 309–309.

General Data Protection Regulation. (27 April 2016). 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) (Text with EEA relevance). *EUR-Lex - 32016R0679* https://eur-lex.europa.eu/eli/reg/2016/679/oj.

HLEGAI. (8 April 2019). High-Level Expert Group on Artificial Intelligence, EU - Ethics guidelines for trustworthy AI. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

HLEGAI. (18 December 2018). High-Level Expert Group on Artificial Intelligence, EU - Draft ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/draft-ethics-guidelines-trustworthy-ai.

HLEGAI. (26 June 2019). High-Level Expert Group on Artificial Intelligence, EU - Policy and investment recommendations for trustworthy artificial intelligence. https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence.

Mokander, J., & Floridi, L. (2021). Ethics-based auditing to develop trustworthy AI. *Minds and Machines* 10.

Mökander, J., Morley, J., Taddeo, M. &, Floridi, L. (forthcoming). Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. *Science and Engineering Ethics*.

Roberts, H., Cowls, J., Hine E., Mazzi, F., Tsamados, A., Taddeo, M. &, Floridi, L. (forthcoming). Achieving a 'Good AI Society': Comparing the aims and progress of the EU and the US.

Roberts, H., Cowls, J., Hine, E., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). China's artificial intelligence strategy: Lessons from the European Union's 'ethics-first' approach. *Available at SSRN 3811034*.

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. *AI & SOCIETY, 36*(1), 59–77.

Scott, J. (2014). Extraterritoriality and territorial extension in EU law. *The American Journal of Comparative Law, 62*(1), 87–126.