weizenbaum journal OF THE DIGITAL SOCIETY

03 SPECIAL ISSUE 03

WEIZENBAUM JOURNAL OF THE DIGITAL SOCIETY
Volume 3 \ Issue 3 \ w3.3.7 \ 10-26-2023
ISSN 2748-5625 \ DOI 10.34669/WI.WJDS/3.3.7

Information on this journal and its funding can be found on its website:
https://wjds.weizenbaum-institut.de

KEYWORDS

artificial intelligence

regulation

artificial intelligence act

VOICES FOR THE NETWORKED SOCIETY

How Should We Regulate AI?

Herbert Zech[1]

Weizenbaum Institute for the Networked Society / Humboldt University of Berlin
herbert.zech@hu-berlin.de

ABSTRACT

In the last decade, artificial intelligence (AI) – which describes the mimicking of human intelligence using technology – has made significant progress. Driven by algorithmic design, computing power and large amounts of training data, machine learning has transformed information technology, which can now augment and replace human intelligence, something that was thought impossible just a decade ago. In 2018, the European Commission labelled AI a transformative technology with the potential to raise new ethical and legal questions.[2] Now, with the advent of generative AI, which can create content that could previously only be created by human beings, this potential has become visible to the wider public. At the same time, the European Commission's proposal for an Artificial Intelligence Act (AIA)[3] (which is now entering the final legislative stage)[4] indicates its intentions to regulate AI. This comment wishes to highlight some key points regarding the regulation of artificial intelligence and, in doing so, comment on the current proposal.

[1] Herbert Zech is a biologist and a legal scholar. He is Full Professor of Civil Law, Technology Law and IT Law at Humboldt University, Berlin, and is a Director of the Weizenbaum Institute for the Networked Society. The author wishes to thank Lea Ossmann-Magiera, LL.M. for valuable input and comments.

[2] European Commission, Communication Artificial Intelligence for Europe, COM(2018) 237 final, p.3.

[3] European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), COM(2021) 206 final.

[4] See https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.

1 Do Not Confuse Artificial with Human Intelligence

The first thing to keep in mind is that, despite their achievements, AI systems are not humanlike in all aspects. Despite being able to mimic and replace human intellectual capabilities, AI is and remains a tool. Where previous technologies provided the means of transforming energy and processing matter, modern AI provides a means of processing information in a manner hitherto impossible. Arguably, the point that information technology has reached represents a turning point comparable to Watt's improvement of the steam engine in the context of energy-harnessing technology.[1] Although generative AI can create humanlike content, this does not make it humanlike as such. Therefore, the seductive idea of treating AI systems as human actors in legal regulations (as in the concept of an electronic person) should be treated with care.[2]

AI represents a new kind of intelligence. Although AI can increasingly perform tasks that could previously only be performed by humans, there remain huge differences.[3] First, AI systems possess the ability to absorb and process huge amounts of information in small amounts of time. Second, AI systems can be connected directly to the internet, to machines and to each other. Finally, AI systems are quite inefficient in terms of energy consumption. Furthermore, while – being machines – they have no moral compass (and, currently, show a tendency to make things up), they may also be simply switched off. Therefore, AI should be treated as a tool (albeit a very powerful one) rather than a person. Despite the debate concerning how to define AI and, with this, the scope of application, the AIA clearly follows this approach.

2 Regulate Only When and as Much as Necessary

When regulating new technologies – whether as products (the apparent approach of the AIA, providing one ignores Title II/Article 5), facilities or services – a risk-based approach is necessary. This means that the risks associated with the new technology are evaluated in advance (the sixth consideration of this comment discusses the limitations of this approach), and regulatory measures are introduced accordingly. Only where sufficiently huge risks exist is interfering with the market necessary and admissible. The smaller the risk,

[1] Brynjolfsson/McAfee, The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies, 2014, Chapter 1.

[2] Zech, in Eifert (ed.), Digitale Disruption und Recht, 2020, p. 29, 42–43.

[3] See e.g. Geoffrey Hinton in an interview with Josh Taylor and Alex Hern, https://www.theguardian.com/technology/2023/may/02/geoffrey-hinton-godfather-of-ai-quits-google-warns-dangers-of-machine-learning.

the less restrictive regulation needs to be.[4] This approach has also been adopted by the European Commission in its focus on high-risk AI systems (Title III/ Articles 6 to 51).

Important considerations include risks to the health and safety or fundamental rights of natural persons. In addition to classical risks to clearly defined interests (e.g. health and property), more amorphous risks concerning personality and non-discrimination appear. This is due to the information-processing nature of AI technology. The explanatory memorandum names features of AI that engender these risks as “opacity, complexity, bias, a certain degree of unpredictability and partially autonomous behaviour of certain AI systems”. Such risks have been discussed in the literature for quite a long time and are definitely worth being addressed.[5]

3 Try to Be Specific

Debate continues around whether the horizontal regulation of AI as a technology – that is, regardless of the sector of deployment – is as advisable as in the case of other technologies (e.g. genetic engineering and nuclear energy technology). Risk-specific regulation is paramount. However, the risks are often difficult to determine. One factor is differences in technology (e.g., artificial neural networks, other machine learning technologies, symbolic AI) and another is the sector in which the system is used (e.g., healthcare, automotive, entertainment).

By definition, AI regulation is technology-specific (hence the list of AI techniques and approaches in Annex I of the original proposal). However, it is difficult to ascertain what can and should be defined as AI and, hence, associated with specific risks. The history of the AIA associates this problem with first concerning the regulation of almost any automated decision-making and then being restricted to autonomous systems. This introduces problems of defining autonomy, which likely refers to machine learning. Arguably, machine learning – with its shift from programming to training – is what makes current AI so simultaneously powerful and risky. Although this is a clear enough concept, whether it can be used as the basis for legal regulation remains to be seen.

[4] Cf. the “criticality pyramid” proposed by the Data Ethics Commission, Opinion of the Data Ethics Commission, 2019, p. 19, https://www.bmj.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN.pdf?__blob=publicationFile&v=2.

[5] See: Wischmeyer, Regulierung Intelligenter Systeme, Archiv des öffentlichen Rechts 143 (2018), 1, 18-65; Mainzer, Künstliche Intelligenz – Wann übernehmen die Maschinen?, 2019, p. 216–226; Martini, Blackbox Algorithmus, 2019, p. 27–64; Zech, Risiken Digitaler Systeme, 2020, p. 26–47; Hacker, The European AI Liability Directives, 2023, p. 57–62.

Moreover, AI will sooner or later be introduced to every aspect of society (much like information technology in general). Just as there is no all-encompassing computer regulation, there will be problems with AI regulation. The answer might involve emphasising sector-specific regulation. Existing sector-specific regulations, such as regulations for medical devices, should be (and are being) updated to be “fit for AI”.

Although the AIA is a horizontal regulation, it addresses these concerns, including an interface to existing regulations within the definition of he high-risk AI systems that are intended for use as a safety component of products. Using an Annex to define the second category of high-risk AI systems (Annex III) helps to simultaneously create legal certainty and maintain flexibility. The annex may also allow for the introduction of not only technology-related but also sector-related aspects.

4 Consider the Whole Value-Chain

Under the new paradigm, a new kind of value chain evolves that resembles the knowledge value chains associated with human innovations. Untrained AI must be trained with data. AI may also be trained in several stages, with pre-trained AI becoming an increasingly important product. Finally, the use of AI either generates useful results (e.g. content) or causes damage. The important role of training data is already being addressed by Article 8 of the AIA, which concerns data and data governance. The debate about general-purpose AI in the later stage of the legislative process demonstrates the increasing awareness of the different steps in the value chain.

5 Consider Risks and Benefits

Although the risks of AI are at the forefront of the current regulatory debate, the relevant legal framework also has to consider the potential benefits (which are considerable). This is reflected in the debates about IP law (as an incentive for innovation) regarding trained and untrained AI systems, AI-generated content and the use of content to train AI. Regulating risks and benefits has to be seen in parallel.[6] However, for benefits, as in the case of risks, the state should only intervene where necessary. Because no market failure is currently observable or imminent, no new IP rights need to be created for the time being.

[6] Zech, Einführung in das Technikrecht, 2021, p. 83 – 84.

6 Consider Limited Knowledge

As in the case of any novel technology, AI systems create knowledge problems for the regulator (known as the Collingridge dilemma[7]). One answer is to include within the regulatory framework elements of flexibility, for example, referring to industry standards. This has been adopted by the AIA (cf. Article 40). However, this has also led to criticism due to the delegation of decisions concerning risk to private actors. An alternative approach sees regulators profit from the technological knowledge of private actors through co-regulation mechanisms.

Yet another strategy prefers indirect regulation (incentives) over direct regulation. In this case, liability for AI must be considered an integral part of the regulatory framework. Unfortunately, the European Parliament's proposal to introduce a strict liability rule for high-risk AI[8] was not adopted by the European Commission. Instead, it was replaced by a proposal focused on easing the burden of proof within existing liability regimes[9]. However, with its proposal for amendments to existing product liability law[10], the European Commission also created a suitable (albeit not strict) liability rule for AI.[11]

7 Transparency Is Important – But Not a Panacea

AI creates several transparency problems. Opacity is considered a main risk source in the AI context (cf. recital 47). Therefore, the AIA ensures "a certain degree of transparency" (ibid.) by creating obligations for the providers of high-risk AI systems (cf. Articles 13, 43, 64) to users, notified bodies (conformity assessment bodies) and market surveillance authorities. In addition, transparency obligations for certain AI systems have been created (e.g. Article 52(1): systems intended to interact with natural persons).

[7] Collingridge, Social Control of Technology, 1980, Chapters 1 and 2.

[8] European Parliament, Regulation of Civil Liability for Artificial Intelligence of 20.10.2020, P9_TA-PROV(2020)0276.

[9] Proposal for a Directive on liability for defective products, COM(2022) 495 final.

[10] Proposal for a Directive on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), COM(2022) 496 final.

[11] Wagner, Liability Rules for the Digital Age - Aiming for the Brussels Effect, 2023, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4320285, p. 4 – 5.

Transparency is also one of the requirements for achieving Trustworthy AI, a concept advanced by the High-Level Expert Group on Artificial Intelligence set up by the European Commission.[12] It is important, especially in the case of AI systems that are inherently opaque (such as artificial neural networks). Introducing transparency requirements as a legal answer to transparency problems may also be considered a less restrictive regulatory measure than authorisation requirements or outright bans. However, user information as a regulatory tool has its limits. As with any complex technology, the effects of such requirements are limited due to the cognitive and behavioural limits of human beings. Therefore, transparency cannot substitute for trust. There are many technologies that we do not understand but use every day because we trust in their safety.

8 Ensure Human Control and Responsibility

Another requirement of Trustworthy AI is human agency and oversight.[13] Humans must keep control of the tool that is AI. This may be achieved by mechanisms such as a human-in-the-loop, human-on-the-loop or human-in-command design. However, because measures must be proportionate and reasonable, they depend on the specific use case and the associated risk. Not all applications work with a human in the loop. The AIA takes this into account (cf. Article 14 on human oversight).

Control is also the basis for attributing responsibility. Responsibility means attribution of the effects (of the development, dissemination and use of AI) under legal rules, whether, for example, regulatory law, liability law, criminal law or contract law. As such, there is also a close link between direct regulation and liability: Ensuring control enables attributing liability. With AI systems (and automated systems in general), it has to be considered that control has shifted from the users to the providers. Accordingly, it is with good reason that the AIA puts the bulk of obligations on the providers and only a small share on the users (Article 29). In liability law, this is reflected in the European Commission’s shift towards product liability (see this comment’s sixth consideration).

[12] High-Level Expert Group on Artificial Intelligence, Ethics guidelines for trustworthy AI, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, p. 14–20.

[13] Ibid.

9 Ensure Backstops

AI may well be taken up by society in a way that it is difficult to turn back from should the need arise. The problem of retrievability is well-known from other technologies and may also be tied to the precautionary principle (cf. Article 191(2) TFEU regarding the environment). How it may be achieved with AI seems less clear. Society should not become dependent on AI. But should it get used to it, then – as in the case of digital technology more generally – there may well be no way back.

This raises the question of a social safety net in the event that regulation fails. Older technologies mainly endanger health, for which there are social health systems. With AIA contemplating risks across a much broader spectrum of interests (see this comment's second consideration), the problem is more complex. One answer might be introducing a social insurance system that specifically addresses AI risks (insurance for certain types of IT-related accidents).[14]

10 Strive Towards an AI-Prepared Society

Human agency, as set out in the ethics guidelines on Trustworthy AI, entails that users are able to make informed autonomous decisions regarding AI systems.[15] This underpins, as a final aspect, the importance of education. The public needs to be prepared for the spread of AI systems by raising awareness, enhancing digital literacy, and creating a better understanding of how AI systems work. This kind of digital education, with the aim of creating "AI-preparedness", goes well beyond the scope of the AIA. It is an important task for all public learning institutions, not least of all the Weizenbaum Institute.

[14] Zech, ERA Forum (2021) 22:147 – 158, at 156.

[15] High-Level Expert Group on Artificial Intelligence, ibid.