

EU AI ACT AND ITS CONDITIONS FOR HUMAN FLOURISHING: A VIRTUE ETHICS PERSPECTIVE

Salla Westerstrand

Turku School of Economics, University of Turku (Finland)

sakrpon@utu.fi

EXTENDED ABSTRACT

Recent developments in Artificial Intelligence (AI) have accelerated initiatives to guide and regulate the use thereof. In the European Union (EU), the European Commission (Commission) gave its proposal for a regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) (hereinafter: EU AI Act) in April 2021 (European Commission 2021). The proposal was followed by intense political discussions and amendments, leading to the adoption of the General Approach of the Council of the EU in November 2022 and amendments adopted by the European Parliament in June 2023. Currently, the draft is in trilogue negotiations between the three EU institutions and is scheduled to be adopted by the end of 2023.

Whereas the upcoming EU AI Act is not an ethics guideline², it is of high relevance when evaluating the societal conditions that guide the ethical directions of AI systems development in Europe but also globally. First, unlike the Ethics Guidelines for Trustworthy AI published by the Commission in 2019 (HLEG, 2019), the EU AI Act is legally binding and introduces sanctions for providers and users of the regulated AI technologies upon non-compliance. Therefore, it can be expected that the EU AI Act contributes to setting the basis for ethical directions of AI development.

Second, it addresses certain phenomena that have been shown to be subject to ethical dilemmas. We could ask, for instance: is it justified to exercise effective surveillance in public spaces and hence limit human freedom to protect citizens' physical safety against a terrorist attack (Almeida et al., 2022; Zuboff, 2019)? Can it be acceptable to deploy highly accurate algorithms in recruitment and thus find better fitting workplaces for most people more efficiently, if that means a higher risk for systematic discrimination of minorities (for a review of ethics of AI in recruitment, see Hunkenschroer & Luetge, 2022)? Both are situations that fall into the scope of the EU AI Act. It includes prohibitions to AI-powered surveillance and deems AI used in recruiting high-risk and thus subject to further requirements. Furthermore, if the EU AI Act leads to the adoption of corresponding measures in non-EU countries similar to the case of the General Data Protection Regulation (GDPR) (a.k.a. *Brussels effect*), it is anticipated to have a global impact on the direction which AI is being developed (Siegmann & Anderljung, 2022). Hence, despite arguably being a Eurocentric perspective and thus not reflecting the full global discourse around AI development, it is expected to offer a fruitful starting point for better

² The Commission gave its recommendations, Ethics Guidelines for Trustworthy AI, in 2019, which is available here: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

understanding the ethical directions of ongoing AI development, mainly in the EU but also beyond.

Still, is compliance with the EU AI Act a guarantee or even a promise of ethical AI? As the latest versions of the EU AI Act were introduced only recently, we still lack understanding of what kind of ethical implications the European regulation could have on AI development when adopted. Understanding the ethical directions implied by the EU AI Act is essential for several reasons: First, it offers the providers of AI systems, namely IS practitioners, a better understanding of what can be achieved by EU AI Act compliance in terms of ethical AI, and what perhaps falls out of the scope of the European legislation and needs to be done by other means. Second, it offers guidance for policymakers to fill the gaps left behind by the EU AI Act in order to work towards more ethical AI systems. Lastly, it offers ground for IS researchers in academia and the private sector to start exploring practical methods and solutions for mitigating ethical issues in AI systems.

Hence, to work towards this deeper understanding and ethical AI systems, we seek a response to the following research question:

What kinds of ethical directions does the EU AI Act imply for AI development?

In this paper, we approach the question from the perspective of virtue ethics, which is a much-discussed branch of moral philosophy originating from the work of Aristotle. We concentrate on the perspective introduced by Bynum (2006), inspired by Norbert Wiener, James Moor, and Luciano Floridi. According to Bynum, only human flourishing can create conditions for ethical action. This perspective of virtue ethics has been shown to be relevant in the context of AI ethics (Stahl, 2021; Stahl et al., 2022; Stenseke, 2021), which makes it a theoretically interesting starting point for analysing the upcoming EU legislation. As Stahl et al. (2022) demonstrate in their study that analyses the role of a European Agency for AI introduced in the EU AI Act, flourishing ethics can serve as a fruitful perspective to shed light on the ethical implications of the upcoming European legislation.

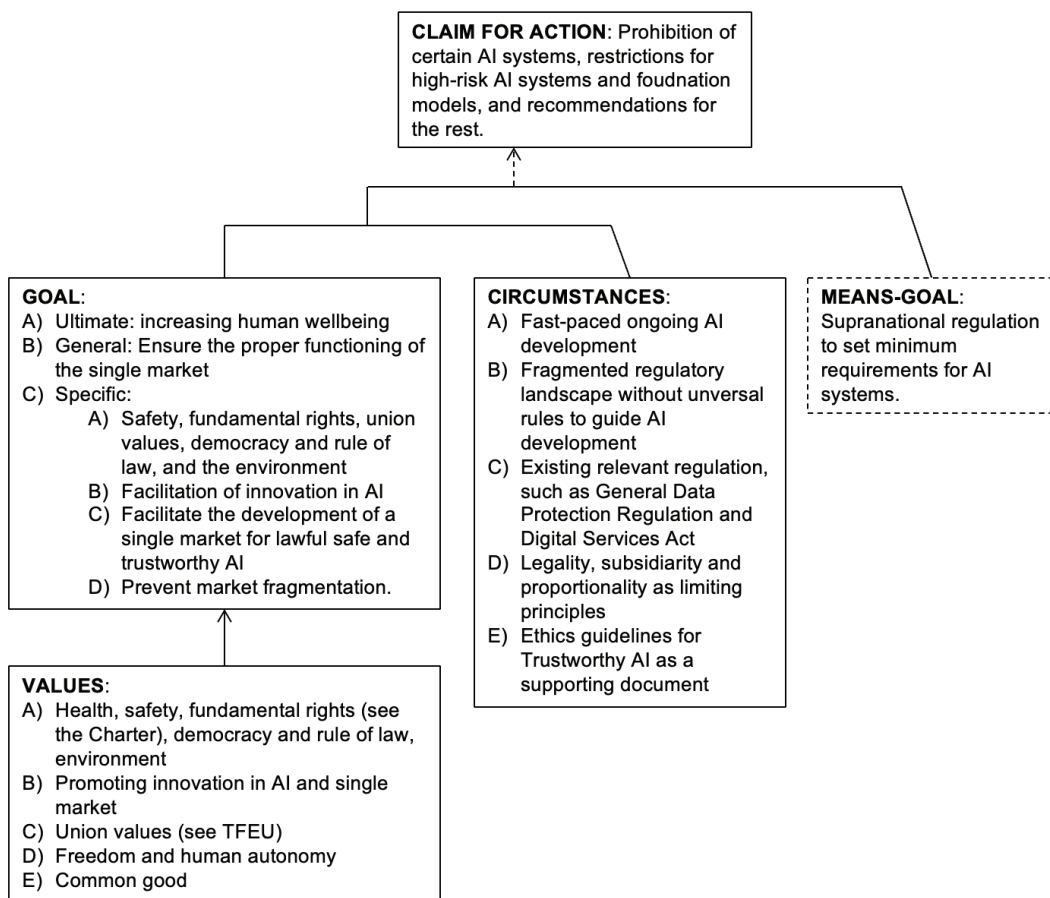
The analysis is conducted using the three latest and major versions of the EU AI Act, i.e., the original proposition by the European Commission, the General Approach adopted by the Council of the European Union, and the amendments adopted by the European Parliament in June 2023, as they are assumed to give the most complete picture available of the main elements of the EU AI Act at the time of writing. The analysis is thus conducted before the results of the trilogue negotiations between the Parliament, Commission, and the EU Council are finalised.

The Act is analysed from the perspective of critical theory, using methodologies of Critical Discourse Studies and Political Discourse Analysis. Critical theories share concern around freedom, autonomy, and human emancipation (Adorno & Horkheimer, 1979), which is often reflected in studies as an interest towards power relations in society (Van Dijk, 2017; Waelen, 2022). They highlight the pragmatic nature of science and knowledge, aiming not only to describe but also to change society by challenging existing paradigms and suggesting alternative courses of action (Delanty & Harris, 2021; Orlikowski & Baroudi, 1991; Stahl, 2008; Waelen, 2022). Approaching ethics in the EU AI Act from this perspective paves the way for justified critique and suggestions for practitioners, policymakers, and deployers of AI technologies, which contributes to directing IS development in a more ethical direction. Such an aim can be considered desirable if not a responsibility for an IS researcher (Chiasson et al., 2018).

Furthermore, we believe that in the age of increasing complexity noted by (Benbya et al., 2020), and the fast-paced development of new systems and user interfaces, offering such guidance helps with ensuring that both AI practitioners and policymakers are informed by scientific research when defining the conditions and technical specifications that shape people’s lives.

We build upon the principles for critical IS research offered by Myers & Klein (2011), and critical discourse studies based on Jürgen Habermas’s work. We structure the ethics discourse in the EU AI Act around the argumentation scheme introduced in Political Discourse Analysis by Fairclough and Fairclough (2013). The argumentation of the EU AI Act can be illustrated in Figure 1.

Figure 1. Argumentation of the EU AI Act, following the structure of practical arguments by Fairclough and Fairclough (2013).



According to Bynum (2006), the central elements for human-centred flourishing ethics derived from Aristotle’s virtue ethics are the following:

- Human flourishing is at the centre of ethics.
- Humans are social, and hence human flourishing requires a connection to society.
- In order to flourish, humans need to do what they are best equipped to do.

- Flourishing requires humans to reason theoretically and practically using intellect and practical judgment. This leads to acting according to the reasoning, in a societal context.
- The key to mastering practical reasoning, and thus to being ethical is "the capacity to deliberate well about one's overall goals and carry out that action."

Analysing the EU AI Act through the lens of these elements demonstrates that the EU AI Act is likely to have impacts on a) human autonomy through practical reasoning, intellect, and deliberation about one's goals; b) connecting with society; and c) doing what humans are best equipped to do, all of which contribute to human flourishing. The present study shows an acute need for further research concerning the ethical implications of AI regulation from several different ethical perspectives. We hope that the results of the present paper encourage researchers, policymakers, and industry professionals to explore the implications of their AI systems that fall outside the mandate and political emphases of the EU legislation in order to strive towards more ethical future AI systems.

KEYWORDS: Artificial Intelligence ethics, virtue ethics, human flourishing, EU AI Act, ethics of AI regulation.

REFERENCES

- Adorno, T. W., & Horkheimer, M. (1979). *Dialectic of Enlightenment*. Verso.
- Almeida, D., Shmarko, K., & Lomas, E. (2022). The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: A comparative analysis of US, EU, and UK regulatory frameworks. *AI and Ethics*, 2(3), 377–387. <https://doi.org/10.1007/s43681-021-00077-w>
- Benbya, H., Nan, N., Tanriverdi, H., & Yoo, Y. (2020). *Complexity and Information Systems Research in the Emerging Digital World* (SSRN Scholarly Paper No. 3539079). <https://papers.ssrn.com/abstract=3539079>
- Bynum, T. W. (2006). Flourishing Ethics. *Ethics and Information Technology*, 8(4), 157–173. <https://doi.org/10.1007/s10676-006-9107-1>
- Chiasson, M., Davidson, E., & Winter, J. (2018). Philosophical foundations for informing the future(S) through IS research. *European Journal of Information Systems*, 27(3), 367–379. <https://doi.org/10.1080/0960085X.2018.1435232>
- Delanty, G., & Harris, N. (2021). Critical theory and the question of technology: The Frankfurt School revisited. *Thesis Eleven*, 166(1), 88–108. <https://doi.org/10.1177/07255136211002055>
- European Commission (2021). Proposal for 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. COM/2021/206 final. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>
- Fairclough, I., & Fairclough, N. (2013). *Political Discourse Analysis: A Method for Advanced Students*. Routledge.

- HLEG. (2019). *Ethics Guidelines for Trustworthy AI*. High-Level Expert Group on AI by the European Commission.
- Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda. *Journal of Business Ethics*, 178(4), 977–1007. <https://doi.org/10.1007/s10551-022-05049-6>
- Myers, M. D., & Klein, H. K. (2011). A Set of Principles for Conducting Critical Research in Information Systems. *MIS Quarterly*, 35(1), 17–36. <https://doi.org/10.2307/23043487>
- Orlikowski, W. J., & Baroudi, J. J. (1991). Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information Systems Research*, 2(1), 1–28. <https://doi.org/10.1287/isre.2.1.1>
- Siegmann, C., & Anderljung, M. (2022). *The Brussels Effect and Artificial Intelligence: How EU regulation will impact the global AI market* (arXiv:2208.12645). <https://doi.org/10.48550/arXiv.2208.12645>
- Stahl, B. C. (2008). The ethical nature of critical research in information systems. *Information Systems Journal*, 18(2), 137–163. <https://doi.org/10.1111/j.1365-2575.2007.00283.x>
- Stahl, B. C. (2021). Concepts of Ethics and Their Application to AI. In B. C. Stahl (Ed.), *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies* (pp. 19–33). Springer International Publishing. https://doi.org/10.1007/978-3-030-69978-9_3
- Stahl, B. C., Rodrigues, R., Santiago, N., & Macnish, K. (2022). A European Agency for Artificial Intelligence: Protecting fundamental rights and ethical values. *Computer Law & Security Review*, 45, 105661. <https://doi.org/10.1016/j.clsr.2022.105661>
- Stenseke, J. (2021). Artificial virtuous agents: From theory to machine implementation. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-021-01325-7>
- Van Dijk, T. A. (2017). *Discourse and Power*. Bloomsbury. <https://www.bloomsbury.com/us/discourse-and-power-9780230574090/>
- Waelen, R. (2022). Why AI Ethics Is a Critical Theory. *Philosophy & Technology*, 35(1). <https://doi.org/10.1007/s13347-022-00507-5>
- Zuboff, S. (2019). *The Age of Surveillance Capitalism*. <https://www.hachettebookgroup.com/titles/shoshana-zuboff/the-age-of-surveillance-capitalism/9781610395694/?lens=publicaffairs>