**Where Law and Ethics Meet: A Systematic Review of Ethics Guidelines and Proposed Legal Frameworks on AI**

Désirée Martin and Michael W. Schmidt,
Karlsruhe Institute of Technology, Germany
*International Conference on Computer Ethics: Philosophical Enquiry (CEPE) 2023, Chicago, IL*

**Extended Abstract**

In the last decades, the application of Artificial Intelligence has come a long way and progressed enormously. However, there are also corresponding risks, insecurities, and uncertainties, which should be considered normatively.

Embedding ethics from the start into the design of AI and throughout its whole lifecycle is directed towards two goals: Firstly, to maximize the benefits for individuals, society, and the life on earth, for example by fostering well-being and enhancing creativity, individuality, dignity, and autonomy. Secondly, to avoid, or at least, to minimize the risks and unintended harmful consequences for individuals, society, living nature, or the planet (High-Level Expert Group on Artificial Intelligence set up by the European Commission, 2019, p. 4). To reach these goals, a variety of AI ethics guidelines has emerged (one example is Floridi et al., 2018).

In addition to a general ethical understanding of AI-related issues, a legal framework is necessary to ensure that ethical values and principles, which are essential in light of basic rights and liberties, are taken into consideration in development and appliance of AI technologies. However, current legislation, e.g., at the level of the European Union, is not in line with the fast developments regarding AI technology. Therefore, the EU Commission proposed the 'Artificial Intelligence Act' (EU Commission, 2021) and the 'EU AI Liability Directive'(EU Commission, 2022). Another blueprint for a legal framework regarding AI technology is currently proposed by the US government (White House Office of Science and Technology Policy, 2022).

From an ethical perspective, there have been efforts to systematize the currently proposed AI ethics guidelines. Reviews like 'The global landscape of AI ethics guidelines '(Jobin et al., 2019), 'The Ethics of AI Ethics. An Evaluation of Guidelines'(Hagendorff, 2020), 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI '(Fjeld et al., 2020), and 'Linking Artificial Intelligence Principles'(Zeng et al., n.d.) offer an overview about commonly proposed values and principles. Nevertheless, none of these papers compares the ethics guidelines with regulatory proposals. Our paper aims to fill this research gap and to deliver a comparison of the main AI ethics guidelines and the currently proposed legal frameworks on AI.

Methodologically, our systematic comparison is based on a literature review of six widely discussed AI ethic guidelines (2017 Asilomar conference (Benficial AI), 2017; Abrassart et al., 2018; Floridi et al., 2018; High-Level Expert Group on Artificial Intelligence set up by the European Commission, 2019; IEEE, 2021; The IEEE Initiative on Ethics of Autonomous and Intelligent Systems, 2017) and five proposed AI legal frameworks and international conventions (EU Commission, 2021, 2022; OECD, 2019; UNESCO, 2022; White House Office of Science and Technology Policy, 2022). We focus on keywords that are understood as principles, values, recommendations or requirements in the texts. On the one hand, some keywords are considered in all texts. On the other hand, not all of them are mentioned in all papers. In addition, many of them are not presented on the same level of importance or not under the same category (e.g., as 'value' or as 'principle').

Considering especially the terms ‘value’ and ‘principle’, it is not always clear what they conceptually mean according to the texts, or how we are supposed to distinguish between them. Thus, it is a philosophical challenge to clarify these concepts in the context of AI ethics guidelines and proposed legal frameworks. We point out reasonable solutions to this problem that enable us to categorize the relevant keywords systematically as values or principles.

Many of the keywords, which can be categorized as values and principles, figure in ethical texts as well as in legal texts. Interpreted and systemized, this points to a possible overlapping consensus regarding normative issues of AI (Rawls, 2001, 2005): ‘autonomy’, ‘explainability’, ‘human rights’, and ‘justice’, for example, are shared ethical values and principles for AI that are also essential from a political perspective and therefore specific regulatory requirements for AI systems might reasonably be connected with them.

Additionally, we identify relations between several values and principles. Just counting the usages of specific keywords cannot illustrate the relation and hierarchical levels between them. It could even lead to false assumptions about significance, if one focuses only on the number of usages. From a systematic perspective, we propose that some principles can reasonably be ordered as higher-level or lower-level ones. ‘Traceability’, for example, may be seen as lower-level principle in relation to the higher-level principle of ‘explainability’. Indeed, most of the keywords are lower-level principles that can be instrumentally assigned to one or more higher-level principles. However, both, higher-level and lower-level principles are in the range of what is typically understood as mid-level principles in applied ethics (cf. Beauchamp & Childress, 2019).

Besides providing a systematic and relational list of moral values and principles, which are widely shared in the moral and legal realm, we also provide explications of these values and principles based on our synoptic findings. This serves to foster a better understanding of these values and principles, which is essential for assessing the acceptability of AI technology and its application.

A more specific as well as important finding of the systematic comparison is that in the current legal proposals, the value of ‘well-being’ and the principle ‘beneficence’ are not yet operationalized. However, well-being and beneficence are shared elements of the main AI ethics guidelines. This knowledge can be useful for the further development of the legal system by illustrating important ethical aspects in AI that are not (yet) transferred to legal guidelines. It is an open question how the legal system should deal with this challenge. Is an implementation missing because well-being and beneficence are hard to operationalize and to measure? If so, are there feasible solutions? Alternatively, are well-being and beneficence missing in the legal proposals, because they are not essential from a political point of view or no proper legal objects? By identifying these urgent questions, our paper contributes to research on the intersection of law and ethics, which is relevant for a responsible societal implementation of AI technology.

**References**

2017 Asilomar conference (Benficial AI). (2017). *Asilomar AI Principles*. Future of Life Institute. https://futureoflife.org/ai-principles/

Abrassart, C., Bengio, Y., Chicoisne, G., de Marcellis-Warin, N., Dilhac, M.-A., Gambs, S., Gautrais, V., Gibert, M., Langlois, L., Laviolette, F., Lehoux, P., Maclure, J., Martel, M., Pineau, J., Railton, P., Régis, C., Tappolet, C., & Voarino, N. (2018). *Montreal Declaration for a Responsible Development of Artificial Intelligence*. Announced at the conclusion of the Forum on the Socially Responsible Development of AI. https://www.montrealdeclaration responsibleai.com/the-declaration

Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics* (Eighth edition). Oxford University Press.

EU Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, COM/2021/206 final,*. https://eur-lex.eu ropa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

EU Commission. (2022). *Proposal for a Directive on adapting non contractual civil liability rules to artificial intelligence*.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3518482

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. https://doi.org/10.1007/s11023-018-9482-5

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

High-Level Expert Group on Artificial Intelligence set up by the European Commission. (2019). *Ethics guidelines for trustworthy AI*. European Commission. https://digital

strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

IEEE. (2021). *IEEE Standard for Transparency of Autonomous Systems*. IEEE Std 7001-2021. https://standards.ieee.org/ieee/7001/6929/

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

OECD. (2019). *Recommendation of the Council on Artificial Intelligence*.

Rawls, J. (2001). *Justice as Fairness: A Restatement* (E. Kelly, Ed.). Harvard University Press.

Rawls, J. (2005). *Political Liberalism: Expanded Edition* (Columbia Classics in Philosophy edition). Columbia University Press.

The IEEE Initiative on Ethics of Autonomous and Intelligent Systems. (2017). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, version 2* (Version 2). https://standards.ieee.org/content/dam/ieee standards/standards/web/documents/other/ead_v2.pdf

UNESCO. (2022). *Recommendation on the Ethics of Artificial Intelligence*.

White House Office of Science and Technology Policy. (2022). *Blueprint for an AI Bill of Rights*. White House Office of Science and Technology Policy.

Zeng, Y., Lu, E., & Huangfu, C. (n.d.). *Linking Artificial Intelligence Principles*. 4.