doi: 10.34190/EAIR.21.023

Comparison of European Commission's Ethical Guidelines for AI to Other Organizational Ethical Guidelines

Minna Nevanperä, Jaakko Helin and Jyri Rajamäki
Laurea University of Applied Sciences, Espoo, Finland
minna.nevanpera@student.laurea.fi
jaakko.helin@student.laurea.fi
jyri.rajamaki@laurea.fi
DOI: 10.34190/EAIR.21.023

Abstract: The European Commission's Ethical Guidelines for AI has a great relevance on the field since it is very thorough. Hagerdorff (2020) has evaluated 22 different ethical guidelines for AI. He found that in 80 per cent of the guidelines handle privacy, fairness and accountability as minimal requirements of responsible AI system. He also noted that these matters in addition to robustness and explainability are more easily solved as technical matters than the social issues that might be arising from the development of AI system. He found that the company codes of ethics were the most minimalistic which was also verified in our paper. Hagerdorff also found that the ethical guidelines usually do not commit to larger societal interests. This paper compares EC's ethical guidelines with those of IBM, Google and IEEE for getting a picture how the ethical issues are approached in commercial environment. The applied analysis method is data-driven; the ethical guidelines are examined and the common themes are noted to appear. The EC's ethical guidelines are the basis of the comparison. Noticed common themes of these guidelines are accountability, transparency and explainability, diversity, inclusion and fairness, safety and security, and societal wellbeing and humanity, even though all the themes are not discussed in all guidelines in detail. It seems that the ethical guidelines usually do not commit to larger societal interests because the societal issues and wider effects that AI has on the society are hard to write on the form of the simple guidelines. The discussion on the effects of the artificial intelligence on the societies needs to be addressed to political decision-makers and wider audience of researchers than just the developers of the AI or business organizations that exploit artificial intelligence. There is also need for involving the users and target groups to this discussion.

Keywords: ethical guidelines for AI, European Commission, ethics of AI

1. Introduction

As artificial intelligence systems become more wide-spread on almost every field, it also raises the importance of considering AI from an ethical stand point. Academic discussion, as well as discussion in other contexts, regarding ethics of AI has been relevant long before AI was even considered something that could be created in any real sense, but has become increasingly so as more and more businesses as well as public institutions are either more dependent on AI powered systems or have completely incorporated AI systems into their operations. This trend has led to companies and institutions to publishing various policies and guidelines concerning AI development which they themselves utilize during their development processes. The guidelines, however, differ greatly and a consensus between the AI developers is still lacking.

As a mean to affect this trend and direct its progress towards a more desirable trajectory, the European Commission's High-Level Expert Group have created their guidelines promoting the development of a trustworthy A. The group intends the guidelines to be used as standards for AI development in the EU area, which would hopefully lead to the developers gaining an advantage on the global market.

In the guidelines featured in this publication four major key areas seem to appear more frequently than others: transparency, fairness, accountability and explainability. Other less frequently discussed topics such as societal wellbeing usually tend to rise in a context where the discussion goes beyond a simple short listing. It is also notable that some of the problematic areas addressed by the EU expert group are not related to AI ethics per se, but are more technical in nature and involve topics such as safety and security. Still, the ethical issues should not be overlooked for they are of great importance to developing anything that can be considered trustworthy.

2. The European Commission's ethical guidelines for AI

The European Commission Ethical Guidelines for Trustworthy AI (referred to as EU guidelines in the text) was published in April 2019 by European Commission High Level Expert Group on Artificial Intelligence. The EU guidelines set the standard for three major aspects throughout the development of the AI system: Legitimacy, Ethics and Reliability. In general, the fundaments for the EU guidelines are based on human rights which

according to the EU Treaties and EU Charter consist of values such as dignity, freedom, equality, solidarity, justice and citizens' rights. (European Commission, 2020; European Commission, 2019.)

The EU guidelines are human-centric and based on the notion of strengthening the shared European values, which include promoting equality, sustainability and fighting climate change. By setting the guidelines so that these values are promoted, the European Commission is also aiming for European AI developers to gain advantage against competing developers from other regions. (European Commission, 2019.) These guidelines have been recognized to have potential to become highly influential since it has already leading to legislation. The fact that the guidelines are linked to fundamental human rights has been recognized to be an advantage of the European Commission's approach. (Smuha, 2020) Charlotte Stix has studied the characteristics of actionable principles for AI. Three elements of the actionable guidelines are "preliminary landscape assessments; (2) multi-stakeholder participation and cross-sectoral feedback; and, (3) mechanisms to support implementation and operationalizability" (Stix, 2021). European Commission's guidelines are fulfilling all three.

The EU guidelines state, that all AI systems should be developed on four ethical principles. First, autonomy of the individual must be respected and freedoms should not be limited. Second, AI must not cause harm and actively prevent it. Third, AI must be fair and both advantages and costs of the AI divided equally. Fourth, AI system should be explicable and its decision-making process transparent. (European Commission, 2019.)

The EU guidelines include seven requirements, which an artificial intelligence must meet to be considered trustworthy. These are human agency and oversight technical robustness and safety, privacy and data protection, transparency, diversity, non-discrimination and fairness, societal and environmental well-being and accountability. (European Commission, 2019.)

In the EU guidelines it is stated, that fundamental rights and human agency must be preserved and human oversight be present throughout the development of an AI for it to be trustworthy. Should there be a risk, that an action initiated by the AI can cause a human rights violation, the risk should be thoroughly assessed and minimized or justified why such action might be necessary. The AI in question must provide users a chance to act as an active agent as well as sufficient information and tools so that the user can understand the AI's decision-making process and question if necessary. As for the human oversight, it can be sufficiently covered by either human-in-the-loop or human-in-command approach. In practice this would require human oversight to be present either throughout the entire development process of the AI and its design or during the activities of an AI, that's already in use, as well as observing the impact of those activities. The oversight should also include the power to decide in any given situation whether the AI should be used or not. (European Commission, 2019.)

The EU guidelines concerning technical robustness and safety set standards for the level of risk avoidance, reliability, accuracy and predictability for the AI. A trustworthy AI should be resilient to cyberattacks and a contingency plan procedure needs to be prepared in advance, which should include either a rule-based mode activated and run by the system automatically during abnormal situations, or an obligatory human interaction to allow the system to continue operating normally. In addition, current GDPR regulations and legislature concern AI development as well and must be taken into account during the development process. (European Commission, 2019.)

As for privacy protection and data governance standards, The EU guidelines dictate that all data collected and accessed by the AI and used for the purposes of AI learning must be well protected with a limited access from outside and the integrity of the data in question be guaranteed. The data should not be used in any manner which could potentially cause harm and the AI should be designed in a way that it can take into account any errors or inaccuracies as well as biases in the data. (European Commission, 2019.)

The EU guidelines include standards for transparency in the AI development process. The AI decision making progress must be clear enough so that it can be understood and the AI should enable complete back tracking and tracing for the entire process of any decision it makes. The AI must never deceive the people interacting with it into believing that they were interacting with an actual human. Rather transparency must be maintained in this aspect as well. (European Commission, 2019.)

Like the standards on data integrity, the EU guidelines also set standards to prevent any unfair biases to be developed into the AI decision making itself. To ensure fairness, the AI should also be developed so it is accessible and can be used by individuals of any specific group without issues. (European Commission, 2019.)

When it comes to societal and environmental wee-being, the EU guidelines state that AI should promote sustainability and responsibility and all sentient beings ought to be considered as stakeholders. During the AI development process environmental issues as well as social ramifications, including political, should be taken into account. (European Commission, 2019.)

Last, the EU guidelines dictate rules for accountability. These include the possibility to audit and evaluate the AI including its algorithms, collected data and development process. Prevention and mitigation of any negative impacts is crucial and there should be a sufficient enough training data to cover all relevant and plausible scenarios. Also, a plan should exist to handle any situation where the before mentioned guidelines could be compromised. (European Commission 2019.)

2.1 Methodology for developing a trustworthy AI according to EU guidelines

The European Commission has introduced a number of methods and advice for AI developers so that their AI can be considered trustworthy according to the EU guidelines. As for technical methods the Commission mentions that the AI architecture should in itself promote AI decision making that is in line with the fore mentioned guidelines and prevent decisions that are not. This can be achieved through white/black listing behaviors or by sense-plan-act -method. Other technical methods include ethics by design -approach, which incorporates ethical thinking into the developing process, explanation methods and testing and validation process. (European Commission, 2019.)

Methodology in accordance with the EU guidelines can also include different types of regulatory mandates such as legislature, code of conduct, standardizations and certifications in both, the development process of an AI and a fully developed, functioning AI system and its actions and decision-making. Other methods include promoting co-operation via education and awareness as well as participation and dialogue with all stakeholders. All these have been gathered into one comprehensive check list for the usage of AI developers. (European Commission, 2019.) One weakness of the European Commission's guidelines might be the broadness of the guidelines. It is not easy to create an easy- to -follow practice for everyday work based on these guidelines. Clearly, European Commission also has a geopolitical agenda when creating the guidelines for the development and use of artificial intelligence. This is a way to safeguard and protect European values and sovereignty, but also a way to show to other actors the global leadership. (Palladino, 2021.)

3. Ethical guidelines of the commercial actors

In comparison to European Commission's guidelines to AI ethics, it is relevant to take the commercial or organizational guidelines of AI ethics into consideration. It is clear to see when searching for AI guidelines, that many companies do not share their guidelines publicly or their public guidelines are on very general level. However, it is important to see what are the ethical objectives that business life values most relevant and what kind of representation of their ethical values the companies display to the public. It should be noted, that some research believe that ethical guidelines are also limiting the discussion on AI ethics. The guidelines might address the discussion only to matters that are stated in the guidelines and the issues outside or new issues emerging will be underestimated or left outside from the discussion. For example, Nicola Palladino states in her article on epistemic communities, that organizational ethical guidelines serve a role of voluntary protocols that include the values, socioeconomic and political beliefs and interest of those who have created the guidelines (Palladino, 2021).

3.1 IBM's everyday ethics for artificial intelligence

IBM has been one of the most open operators sharing their ethical guidelines. They have published their guidelines in a form of guidebook Everyday ethics for artificial intelligence. The structure of this guidebook is quite similar to European Commission's guidelines. In the beginning they give five areas of ethical focus: Accountability, value alignment, explainability, fairness and user data rights. These are similar to European Commission's most important guidelines and these are more straightforward to take into action. Throughout the guidebook IBM gives use cases and examples, how these ethical values can be executed and they often give

recommended action to take and like European Commission, IBM uses questions for the team to promote ethical discussion. In general, IBM has taken their guidelines well to practical level. (IBM, 2019.)

3.2 Google's artificial intelligence at Google: Our principles

Another big player in a field of AI is Google. Google has published their common level principles for ethical AI. Their emphasis is on general level guidelines. Google's first principle is "Be socially beneficial". This principle is missing on IBM's guidelines, but it is one of the most important guidelines in European Commission's paper. This principle is not easy to take on the practical level since many innovations can be used both good and bad. Google has taken this into account in their principles. They have created a checklist that they use in evaluation of the AI application. These are 1) primary purpose and use, 2) natura and uniqueness, 3) scale and 4) nature of Google's involvement. Google's approach to ethical principles also differ traditional approaches since it gives a guideline to AI applications that Google does not pursue. This includes applications that cause harm (this is also present in European Commission's guidelines) and weaponry. Google also has raised scientific excellence as one of their principles. This means that Google will promote sharing AI knowledge by educational materials, best practices and research. This is aligned with European Commission's goal to share AI knowledge openly. Otherwise, Google's AI principles are very much alike with IBM's and other companies published guidelines including accountability, unfair bias and other common features of responsible artificial intelligence. (Google, 2020).

3.3 IEEE's ethically aligned design

The Institute of Electrical and Electronics Engineers (IEEE) has published ethical guidelines 2019. This guidebook is referred in many company policies considering AI ethics among European Commission's guidelines. The approach to ethics is very similar to European Commission's, but somehow more complex. The basis (pillars) for their recommendations is in human rights, self-determination of data agency and technical dependability. The same way than in European Commission's paper, also IEEE has set general principles to ethically sustainable design. The IEEE's principles are equivalent to European Commissions seven requirements, even though IEEE has eight principles instead of seven. The IEEE's principles are human rights, well-being, data agency, effectiveness, transparency, accountability, awareness of misuse and competence. As we can see the principles are named differently equivalent to European Commission's, but the content is mostly the same. Only notable difference is that awareness of misuse is separate principle (IEEE, 2019; European Commission, 2019).

The content of European Commission's requirements for AI and IEEE's principles are almost identical. Also, how these requirements and principles are mapped to upper-level ethical basis is similar and have the same values included. However, how the principles or requirements are considered to be introduced to action differ. European Commission has introduced an assessment list which has listed a number of questions that AI system should be assessed to be trustworthy. IEEE has introduced issues and recommendations to take ethical guidelines into action. IEEE's approach gives broader recommendations than European Commission's list. Their recommendations are much more on the same level than European Commission's requirements. European Commission's list takes the practical side a step further by giving simple yes and no-questions to developers of AI to consider (IEEE, 2019; European Commission, 2019). However, this EC's detailed checklist might be in some ways limiting and inhibit further ethical thinking.

4. Discussion on the guidelines

Thilo Hagerdorff in 2020 article evaluated 22 different ethical guidelines for AI. He found that in 80 per cent of the guidelines handle privacy, fairness and accountability as minimal requirements of responsible AI system. He also noted that these matters in addition to robustness and explainability are more easily solved as technical matters than the social issues that might be arising from the development of AI system. Hagerdorff also found that the company codes of ethics were the most minimalistic which was also verified in our review. He also found that the ethical guidelines usually do not commit to larger societal interests. This is partly because the societal issues and wider effects that AI has on the society are hard to write on the form of the simple guidelines. (Hagerdorff, 2020.) Larger societal interests are widely present only in European Commission's and IEEE's guidelines. This is due to the nature of these organizations. Their role is more on societal level than single company's. When discussing European Commission, the level is even regulatory and legislative which can influence how the artificial intelligence can be developed and used by the companies and organizations. However, if we consider what are the most influential actors in everyday lives of the users, then we see that the companies like Google and Microsoft become more relevant. It is noted that the analysis of the commercial

actors' guidelines here is rather concise. Broader variety of organizational guidelines for the analysis would be appropriate and the deeper analysis of the values behind the guidelines is needed.

McNamara, Smith and Murphy-Hill reviewed how the ethical guidelines effect the work of the software engineers. They concluded that the ethical guidelines given had almost zero effect on the practices of the professionals. Unfortunately, the study did not examine the reasons why there was no effect. (McNamara et al, 2018). Katie Shilton in her study brought into light that the developers were aware of the ethical issues that might be related to the system they were developing, but they felt that the handling those ethical issues were not their job. (Shilton, 2012.) It has been also discussed that checklists like the one provided by the European Commission, might instruct the development to focus only to the matters that are on the checklist, not the problems that might be completely new or that was not added on the checklist at hand.

5. Conclusions

The guidelines like European Commission's guidelines for trustworthy AI are important since they are good way to create standard procedures for development and use of artificial intelligence. Sharing the best practices and technical solutions for ethically behaving AI are suited for creating standards and policies that are less rigor than binding legislation. This enables flexibility for future solutions that we are not yet able to foresee. The guidelines promote self-regulation of the field.

It must be said, that legislation is to the point a good way to regulate the development and use of AI. For example, GDPR has given boundaries by obligation to data security and privacy, but it is important to recognize that the legislation might be also limiting. It is also important to notice that it is crucial to concentrate on solving the ethical issues where there is no legislation or regulation.

It can be argued whether the ethical guidelines are the best way to produce ethical awareness. It might be better to have some sort of combination of ethical guidelines, ethical training and promoting ethical competence by ethics rounds or some other methods that use real cases that arise from the work of the developers. Also, the technical methods like ethics-by-design or values-by-design are important when considering making artificial intelligence to act responsibly. One subject for the further studies could be to study the effectiveness of the ethical guidelines or to compare different methods of promoting ethical awareness.

In general, the guidelines for artificial intelligence can be seen as an effort from societies, organizations and companies to show that they have taken some time and effort to discuss how the artificial intelligence can affect human life. For commercial partners the public guidelines might differ from the internal ones, but from the public guidelines it is easy to see that the matters discussed are quite similar in all companies with only minor variations. European Commission's guidelines are broader, but it is partly due to the nature of European Commission's role as a European wide governmental organization and legislative actor. European Union has also been very active promoting global ethical discussion on the regulation for the use of artificial intelligence. As an indication of this European Commission has published not only the guidelines for the development and use of artificial intelligence, but In April 2021 European Commission suggests European Union to have new Artificial Intelligence Act. Globally this would be the first law for AI regulation.

Even though some research show that ethical guidelines have not always promoted ethical action, still the guidelines might be good way to promote discussion on the ethics. The guidelines combined with education or promoting ethical competence in other ways might lead more ethical behavior in situations where it is needed. When discussing ethical guidelines of the commercial actors and even though the public ethical guidelines are there more to show to the public that the company has had some thought on the matter than to guide ethical behavior as such, publicly published guidelines might lead the public to have some regulative effect towards the company behavior. However, there is not enough research on the true effectiveness of ethical guidelines or comparison to other methods of promoting ethical awareness. When discussing artificial intelligence, it is important also to discuss the special features of the technology and the technical measures that promote ethical behavior of the technology itself.

Moving forward, as the use of AI becomes more common and applications vary, it is meaningful to shift the focus of the research to one single field at a time and study the effects of AI on that field more closely. The ethical guidelines for healthcare robotics have not been discussed thoroughly and it is a field of study where the

application of the guidelines remains to be seen. The focus of this study will move towards that area and in the next step various stakeholders affected by the development of AI in healthcare robotics should be included in the study. Healthcare robotics are expected to become more relevant in the future and research for the ethical guidelines surrounding the development of the robotics and AI systems directing them should be discussed in advance so that in the future they can be utilized in a safe and ethical manner.

References

European Commission's High-Level Expert Group on Artificial Intelligence (2019) *ETHICS GUIDELINES FOR TRUSTWORTHY AI.*

European Commission (2020). https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en. Referred 19.1.2021.

Google (2020) *Artificial Intelligence at Google: Our Principles.* https://ai.google/principles/ . Referred 28.4.2020.

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

IBM: Everyday Ethics for Artificial Intelligence, IBM design program office. 2019. https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf Referred 27.4.2020

IEEE (2019) *ETHICALLY ALIGNED DESIGN A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems.* Referred 27.4.2020

McNamara, A., Smith, J., Murphy-Hill, E. (2018). "Does ACM's code of ethics change ethical decision making in software development?" In G. T. Leavens, A. Garcia, C. S. Păsăreanu (Eds.) *Proceedings of the 2018 26th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering*—ESEC/FSE (pp. 1–7). 2018. New York: ACM Press.

Palladino, N. (2021) "The role of epistemic communities in the "constitutionalization" of internet governance: The example of the European Commission High-Level Expert Group on Artificial Intelligence". *Telecommunications policy*, 45(6), doi:10.1016/j.telpol.2021.102149.

Shilton, K. (2013) "Values Levers: Building Ethics into Design". *Science, Technology, & Human Values*, 38(3), pp. 374-397. DOI 10.1177/0162243912436985

Smuha, N. (2019) "The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence". *Computer Law Review International*, 20(4), pp. 97-106. doi:10.9785/cri-2019-200402

Stix, C. (2021). "Actionable Principles for Artificial Intelligence Policy: Three Pathways." *Science and Engineering Ethics*, 27(1). doi:10.1007/s11948-020-00277-3