

Position Paper

Consultation on „Ethics Guidelines for Trustworthy AI“ of the European Commission’s High-Level Expert Group on Artificial Intelligence

2019-01-31

Page 1

1. Introduction

As Europe's largest digital association, Bitkom endorses the published draft of the AI Ethics Guidelines and welcomes that the issue is being addressed in an interdisciplinary manner at the European level. We especially support the Commission’s engagement to interact with a broader set of stakeholders in order to develop these guidelines and to share information on the Group's and the Commission's work. The paper represents an important and valuable approach to specify how concrete ethical values can be operationalised in the social, political and economic context of AI. Europe’s ethical values should not only be implemented in the development of AI but also facilitate socio-economic progress. As such, digital ethics can represent a significant competitive advantage within the field of AI.

Bitkom wishes to emphasize that ethical guidelines should be sharply separated from legal issues. Our understanding is that the guidelines are intended to contribute to leveraging this potential. We understand that the guidelines neither constitute nor directly prepare new regulation regarding AI. A tightened legal framework would be detrimental to the European AI ecosystem and thereby constitute a societal disadvantage.

2. General Comments

Bitkom welcomes the two components of Trustworthy AI, stating that it should ensure an ethical purpose (e.g. to respect fundamental rights) but also be *“technically robust and reliable since, even with good intentions, a lack of technological mastery can cause*

Bitkom
Bundesverband
Informationswirtschaft,
Telekommunikation
und neue Medien e.V.
(Federal Association
for Information Technology,
Telecommunications and
New Media)

Kristin Strauch
Artificial Intelligence
P +49 30 27576-133
k.strauch@bitkom.org

Albrechtstraße 10
10117 Berlin
Germany

President
Achim Berg

CEO
Dr. Bernhard Rohleder

Position Paper

Consultation on „Ethics Guidelines for Trustworthy AI“ of the European Commission's High-Level Expert Group on Artificial Intelligence

Page 2|6

unintentional harm” (p. i; 6f).

The published guidelines “are not meant to stifle AI innovation in Europe, but instead aim to use ethics as inspiration to develop a unique brand of AI, one that aims at protecting and benefiting both individuals and the common good. This allows Europe to position itself as a leader in cutting-edge, secure and ethical AI”. (p. ii) We strongly support the idea of protecting the wider public's interest with regard to upcoming trustworthy AI products and services. We thus suggest that the draft includes clear language so that the proposed guidelines can be expected to achieve relevancy and protect the interests and rights of European citizens. The ambition should be to bring the guidelines fully into practice as/via domain-specific ethics code(s). The focus of policy makers should be to strengthen the European AI ecosystem.

In our opinion, the guidelines should also contribute to an understanding of the social learning process regarding AI and increase general trust in AI. As a new basic technology, AI creates a learning process for all stakeholders. In this process, different learning abilities, the willingness to learn and responsibilities have to be taken into account. This certainly also involves consumers, which must also be given the opportunity to partake in this process. In the guidelines draft, however, the consumer mostly appears as an object to be protected (this is due to the “*rights approach to be protected*”) and thus not as a self-empowered subject. This potentially neglects consumer's role within the ethics of AI.

We also want to state that it could be instructive to deal with the ethical aspects of AI on the basis of time scales. For example, the question should be considered whether AI systems should assume full control and decision-making autonomy beyond human capabilities (e.g. under 500 milliseconds) below a certain time. What about hours or days when autonomous decisions can become reversible as more information becomes available? Just because there is a lot of time to reverse a decision, there may still be areas where we do not want the AI to make such decisions without human supervision.

We would also like to draw attention to a general discussion which also concerns the meta-aspects of these guidelines: Within philosophical discourses on technology, the question is discussed to what extent algorithms can actually make decisions. The way we talk about AI has significant impact on this principle and should be taken into account by the guidelines. We propose to speak of “*AI Processes*” and not “*AI Decision Making*” or “*AI Decisions*” as these are key phrases within critics of AI. A more nuanced wording of AI concepts could possibly lead to a higher acceptance of AI by the public.

Position Paper

Consultation on „Ethics Guidelines for Trustworthy AI“ of the European Commission’s High-Level Expert Group on Artificial Intelligence

Page 3|6

3. Rationale and Foresight of the Guidelines

We welcome the fact that the guidelines should not be seen as an end point, but rather as the beginning of a new debate on Trustworthy AI (p. 2). We furthermore like to emphasize the importance of national discussions (e.g. the Data Ethics Commission Ethic of the German Federal Government) on Trustworthy AI, which should also be facilitated and considered by the paper.

The paper also states that the final version of the document will set out a mechanism that enables all stakeholders to formally endorse and sign up to the guidelines “*on a voluntary basis*” (p. 2). This aspect raises important questions such as: If a stakeholder “*formally endorses*” the guidelines, would they substitute already initiated self-binding Codes of Conducts? If a stakeholder would not formally endorse the guidelines, would that create the impression that the stakeholder does not support ethical aspects of AI? And what role do national associations and their member companies play?

On page 5, freedom occurs only in the limited form of “democratic freedom” and is quite insignificant overall. The idea of freedom aims at the responsibility of individuals and organizations. Orientation towards this value would therefore facilitate and promote individual and collective responsibility. The broad use of AI could help in this process, but poses also challenges in regards to avoiding responsibility and self-empowerment. This is a relevant challenge, which should be taken into account.

4. Chapter I: Respecting Fundamental Rights, Principles and Values – Ethical Purpose

At numerous places, terms like „*wellbeing*“ and „*the common good*“ (p. 5) are being addressed without further interpretation. In order to avoid uncertainty, we suggest that the authors should ensure that these terms relate to civil rights. Moreover, we would like to note that the section “*Fundamental Rights of Human Beings*” (p. 6f) does not address the important questions about value assessment and value conflicts.

The paper stresses that in case of harm, AI systems should provide users “*with effective redress*” (p. 10). It should be clarified whether humans are eventually responsible for all potential damage that is caused by an AI system. Operators of an AI system should define

Position Paper

Consultation on „Ethics Guidelines for Trustworthy AI“ of the European Commission’s High-Level Expert Group on Artificial Intelligence

Page 4|6

who is ultimately responsible, as a diffusion of responsibility could trigger distrust among users and thus stifle the overall acceptance of AI systems.

In our view, the guidelines require further concretization regarding the monitoring and willingness to assume responsibility of AI systems during their whole life cycle. Not only the development, but also the everyday deployment of AI systems demand an accountable person - or group of persons - in charge for the processes.

Regarding covert AI systems the draft states that *“(AI) developers should therefore ensure that humans are made aware of – or able to request and validate the fact that – they interact with an AI identity”* (p. 11). The relevance of this topic is exemplified by the ever-increasing use of chat bots (in either written or vocal communication) where it is not always obvious for the user that the communication partner is not a human. Including the obligation to identify such non-human communication partners into a regulation might be helpful. This could be ensured by labelling, for example, chatbots (as distinct from human assistance, see p. 9 and section 5. 2) and by providing alternative formats of communication if exchange/contact with AIs is not desired.

5. Chapter II: Realising Trustworthy AI

In general, we would like to suggest the consideration that the authors differentiate between B2B-based AI systems that are built for a professional context (i.e. flight controller, accounting, social security) and B2C-based AI systems for consumers (i.e. dating apps). The ethical framework and the requirements of Trustworthy AI (cf. p. 13) may differ significantly.

Regarding the accountability of AI system the paper states that *“good AI governance should include accountability mechanisms, which could be very diverse in choice depending on the goals. Mechanisms can range from monetary compensation (no-fault insurance) to fault finding, to reconciliation without monetary compensations. The choice of accountability mechanisms may also depend on the nature and weight of the activity as well as the level of autonomy at play”* (p. 14). This section should also discuss how to handle accountability in the case of severe wrong decisions, e.g. such that cause the loss of human life (e.g. in Health, Autonomous Driving etc.). However, at this point, we would like to stress once

Position Paper

Consultation on „Ethics Guidelines for Trustworthy AI“ of the European Commission’s High-Level Expert Group on Artificial Intelligence

Page 5|6

again the importance of differentiating terms of liability from terms of responsibility more clearly.

The section “*Data Governance*” (p. 14) describes the technical process and challenges of machine learning training, but not how to deal with data acquisition. Since AI is a data-driven model, the ultimate decision is who has the most and highest quality data. This in turn favours populations with a high population density or extensive data collection - the more (high quality?) data the better the AI algorithms. Bilateral agreements are necessary and a kind of data hub that functions as a control instance for the exchange of data between nations.

The paper also states that “*When data is gathered from human behaviour, it may contain misjudgement, errors and mistakes. In large enough data sets, these will be diluted since correct actions usually overrun the errors, yet a trace of thereof remains in the data*” (p. 14f). However, this might be too optimistic. One cannot rely on self-correction due to large enough data sets. Even a large set of data may contain a structural bias which can eventually be passed on to the products (bots, algorithms etc.) that are built from them.

We would like to question the published statement that “*systems should be designed in a way that allows all citizens to use the products or services*” (p. 15). It is already the fact and also highly likely that future AI products and features may appeal in particular to a restricted target group. A “One-design-fits-all”-approach does not seem to be very practical.

In order to address the requirements to achieve Trustworthy AI, we would like to add another non-technical method (p. 18ff) to be employed within the development process: AI systems should come with a clear description of their limits, including the areas they are intended for and those they are not intended for, as well as a description of input data that the system cannot properly cope with (e.g. an animal recognition system that has been trained with data on mammals might not suit well for identifying insects).

In our view, the distinction between “*ethical purposes*” and “*technically robust and reliable*” is necessary. However, the Draft does not clearly distinguish between those two aspects, especially from Chapter II onwards. However, this would make sense. The aspects “*technically robust and reliable*”, to which sections 8 (robustness) and 9 (security) are most likely to be assigned, should be expanded.

Position Paper**Consultation on „Ethics Guidelines for Trustworthy AI“ of the European Commission’s High-Level Expert Group on Artificial Intelligence**

Page 6|6

Regarding “Figure 3: Realising Trustworthy AI throughout the entire life cycle of the system” (p. 18): Please put ‘analysis’ in the top left corner of the four boxes that indicate the recursive flow of actions to maintain and improve over the life cycles - it seems more natural in the mode of reading top to bottom, left to right, that analysis is the start.

6. Chapter III: Assessing Trustworthy AI

It should be emphasised that the steps presented in the section “Assessing Trustworthy AI” can only be the first interventions on the road to create a Trustworthy AI. Their concretisation and implementation requires entrepreneurial initiative in order to implement them and to turn them into economic reality. We thus suggest to stress that organizational impact needs to be taken into account not only during design time, but also during runtime of AI systems.

We would also like to point out that the list of questions currently appears to us to be provisional in some respects. The published questions should be - in order to refer to the introductory chapter of this document - seen as the starting point, not as the end of a discussion on the ethical aspects of AI.

Bitkom represents more than 2,600 companies of the digital economy, including 1,800 direct members. Through IT- and communication services alone, our members generate a domestic annual turnover of 190 billion Euros, including 50 billion Euros in exports. The members of Bitkom employ more than 2 million people in Germany. Among these members are 1,000 small and medium-sized businesses, over 500 startups and almost all global players. They offer a wide range of software technologies, IT-services, and telecommunications or internet services, produce hardware and consumer electronics, operate in the digital media sector or are in other ways affiliated with the digital economy. 80 percent of the members’ headquarters are located in Germany with an additional 8 percent both in the EU and the USA, as well as 4 percent in other regions of the world. Bitkom promotes the digital transformation of the German economy, as well as of German society at large, enabling citizens to benefit from digitalisation. A strong European digital policy and a fully integrated digital single market are at the heart of Bitkom’s concerns, as well as establishing Germany as a key driver of digital change in Europe and globally.