



# CCBE Comments on the Stakeholders' Consultation on Draft Artificial Intelligence Ethics Guidelines

01/02/2019

The Council of Bars and Law Societies of Europe (CCBE) represents the bars and law societies of 32 member countries and 13 further associate and observer countries, and through them more than 1 million European lawyers.

The CCBE appreciates the opportunity to take part of the ongoing “Stakeholders' Consultation on Draft AI Ethics Guidelines”, being a member of the European AI Alliance, and has carefully taken into consideration the draft paper as a result of the discussion currently taking place between the 52 experts of the High Level Expert Group (HLEG). The issues and principles set out in this paper are all significant aspects to consider around the use and development of Artificial Intelligence (AI) systems.

Due to the tight time schedule of the consultation, the comments submitted below are general remarks for the purpose of presenting a preliminary analysis of the issues set out in the Draft Ethics Guidelines for Trustworthy AI and are therefore still subject to any position being developed at a later stage by the CCBE on this topic.

## Introduction: Rationale and Foresight of the Guidelines

- We understand that the HLEG paper (hereafter ‘this paper’) is a starting point for the discussion on a “Trustworthy AI made in Europe”. As a general remark, we appreciate that the experts of the HLEG have provided a comprehensive view on the way how to achieve a “Trustworthy AI” by defining an ethical framework for achieving it.
- The authors not only propose a wish list, but also methods for effectively achieving the established goals. The list of methods includes both technical characteristics and non-technical suggestions. The work done is very sound, but it will remain quite largely theoretical unless a comprehensive case study assessment will complete it. We believe that the final version of the paper should include examples of implementation, on a case-by-case basis, in specific areas. We also note that among the four particular use cases of AI that the final version of the HLEG Guidelines will develop, the case of “Profiling and law enforcement” would be included. Moreover, the CCBE suggests to also include a case scenario on the use of AI in justice, for example in a trial proceeding.
- **The CCBE considers that there is a strong need for having a special discussion on the use of AI in the justice systems.**

As indicated in page 3, the authors acknowledge that, while the scope of those guidelines covers AI applications in general, different situations raise different challenges and thus, a tailored approach is needed given AI’s context specificity.

We understand that the scope of the paper is to provide general overarching principles, but by not considering the specifics of justice systems, we fear that some important issues around the use of AI and automation systems in the field of justice will not be tackled. Since justice plays such a large and special role in the society, we believe that this should be specifically discussed in this paper. In many cases the standards of human behavior are being created (or at least directly applied) in the judicial process, the special use of AI in judicial systems should be more explored in this regard. This discussion could be added in the part 5 of Chapter I – Critical concerns raised by AI.

In every way, the right to a fair trial is one of the fundamental rights, which is the basis for the “Trustworthy AI”, and trustworthy justice is part of the rule of law and principle of democracy.

## **Chapter I: Respecting Fundamental Rights, Principles and Values - Ethical Purpose**

Chapter I deals with ensuring AI’s ethical purpose, by setting out the fundamental rights, principles and values that it should comply with.

- This part focused on concerns related to fundamental rights, individual freedoms, the common good, the environment and the future of humanity. The issues surrounding the development of science and ethics are nothing new. There is indeed a very abundant literature exploring that dimension and the issues arising in that respect. It could be suggested to further complement this paper by integrating those reflections and make it clearer how the reflection we are facing is novel and differ from the issues encountered in other fields of knowledge.
- In any case, the wish that an ethical reflection should go together with the development of AI and intelligent systems which should be designed with respect of fundamental rights, is something the CCBE strongly believes is desirable and necessary.

## **Chapter II: Realising Trustworthy AI**

- This paper does not consider an essential component when applying fundamental rights and moral values, i.e. the possible conflicts and contradictions that may arise between them. We know that ethical problems arise when several fundamental principles are competing or in contradiction with each other. Pretending to build an AI in “compliance with Ethics” would presuppose that there will be a list of existing solutions available to solve all the possible conflicts. This is not the case. The CCBE considers that the paper omits developing these concerns and the inclusion of a note on the possible options how to deal with conflicting fundamental rights or moral values would be highly desirable.
- Another important consideration concerns the question whether an AI system can be entrusted with the task of determining what is right or wrong in a given situation and to what extent we can consider to delegate ethical choices to an AI system? In this context, it is necessary to consider whether we accept the very principle of a delegation of powers to AI allowing it to settle a conflict between values or between the divergent interests and rights of several human beings. Would we accept that an AI determines solutions to ethical or moral problems, with the risk of causing harm because of its choices? This also relates to the question of liability and accountability of AI systems.
- Technology and AI systems are reshaping the decision-making process in both public and private sectors and therefore have also the ability to reshape the relationship between decision-makers and citizens. Any conclusion in this respect should clearly state that it is not proven that the benefits of AI are greater than the risks. There can be some benefits from the perspective of public authorities to make use of such systems for reasons like efficiency or reduction of costs, or from the perspective of citizens: more impartiality (humans vs. machines); equality, legal certainty, and consistency through automated decision-making. However, ensuring that systems comply with the

rule of law is not apparent and this aspect should be carefully considered, especially when the use of automated systems may endanger the principle of procedural rights.

- Recent experience demonstrates that transformation of paper-based processes to electronic ones sometimes resulted in transferring administrative burdens to citizens. In the public sector, these transformations resulted in both cost savings (decrease in the administrative staff required) and faster, more effective processes on the government side. However, such transformations generated further costs on the citizen's side (including their representatives), e.g. new integration costs that were not present in paper-based processes, or IT security costs due to new threats. Citizens may have saved on the post € 10, but have to spend € 100 on IT security and updates. With regard to further transformation of such processes to be able to make use of the increased capabilities provided by AI, it is important to take a look at the context of the processes transformed, and to inspect any unintended effects such transformations may have. These unintended effects are often not cost related, but result in indirect loss of importance of certain values, as collateral damage, such as confidentiality of communications or privacy. We suggest that with regard to transformation of processes in the public sector using AI, this requirement of investigating the unintended effect on context should also be included.
- Similar to the problem mentioned in the previous point, the requirements in Chapter II are often in contradiction with each other. E.g. transparency and safety often require features that result in weakening privacy requirements. If someone designs a system for users with diverse and different disabilities, this might also result in unintentional discrimination for technical reasons. Even with the best intention and with an enthusiasm for transparency, a designer of an AI system could get into a conflict where the trained features show existing discrimination, and the designer has to intervene to avoid further strengthening the discrimination present in the society. We think it could be useful to mention either in Chapter II or Chapter III that in assessing a Trustworthy AI, one also has to address all the requirements identified (even if the list is not exhaustive), and make a human decision on priorities.
- From the paper (see page 7 point 3.3), it is clear that the authors consider the use of AI systems in judicial systems in a way that allows AI to take decisions. Perhaps, this could be the case in some straightforward decision-making process when there is no doubt about applicable moral standards and the way they should be applied – giving the possibility of a human review. An example of a parking ticket violation can be provided in this context. This will be different in complex cases when values are to be applied differently or interpreted in a new way. In these cases, AI should play a supportive role only. This again underlines the need to more carefully consider the topic of the use of AI in the justice field.
- Also, the paper seems to focus on AI-human relations and does not really consider the relations between two AI systems that can influence humans.
- The human-centric conception is a leading principle of the paper. In this respect, it should be considered to state – to avoid any ambiguity – that the ability of artificial intelligence systems to make decisions autonomous from human control, requires specific attention, especially when these decisions can change the legal position of an individual or entity and imply making choices between concurring or conflicting values. In this respect, such autonomous process should be even more carefully assessed, from an ethical point of view, before being introduced in specific fields (such as Judiciary and legal services).
- Humans are the key element in this paper. However, the CCBE wonders whether the effect of AI behaviour towards humanity (as something broader than a human) should not be explored as well?
- The question of ‘transparency’ also encompasses the question of who ‘owns’ an algorithm, e.g. an individual developer, a multinational company with a wide-spread business model or even a state-owned enterprise. This aspect should be reflected more in the paper.
- Whilst the problem of errors is recognised, a discussion on how to remedy such ‘malfunctions’ is missing.

- It might be necessary to conduct a more in-depth analysis of the potential impact of AI with regard to the rights protected by the EU Charter of Fundamental Rights instead of only making reference to “fundamental rights”.

### Chapter III: Assessing Trustworthy AI

Overall, having the capability to generate tremendous benefits for individuals and society, AI also gives rise to certain risks that should be properly managed. We must ensure to follow the road that maximises the benefits of AI while minimising its risks. A human-centric approach to AI should consider to keep in mind that the development and use of AI should not be seen as an end in itself, but as a means to increase human well-being. In this perspective, trustworthy AI is a qualitative factor.

As noted above, the CCBE believes that specific consideration for the use and development of AI in the justice field in light of its potentialities should be separately developed in this paper.

For example, the following important issues and questions may arise in this regard:

- The judicial system is currently in charge of producing solutions to conflicts of norms. It is the role of the judicial system to provide individualized solutions to the ever-present conflicts between principles and ethical values. If the decision is made to let the AI develop its own solutions, then it is a transfer of a priori responsibility for moral choices from the judge to the machine. Here we have again the question of fundamental rights. Is it compatible with human dignity that machines judge men?
- In practice, it would first and foremost be a transfer of responsibility for deciding, from the judge to the designer of the AI. Can such a transfer be considered if the ex-ante guarantees provided by the AI, at the design stage, are not of the same level as those existing in the judicial system? Should the AI designer be independent and impartial, as the judge should be? At least in the event of an AI being implemented to assist (replace?) the judge, this seems to be a necessity.
- Is it otherwise possible to be satisfied with a simple ex post control, in the form of compensation for errors made by the AI, or a right of appeal from the AI's decisions?
- The judicial context also gives rise to other specificities:
  - It can never be assumed that the AI systems implemented effectively respect the principles governing the functioning of AI. The right to review must in all cases be open to the parties concerned.
  - Lawyers should be given the opportunity to verify the compliance of the systems used with the principles identified.
  - Justice is an area in which transparency and accountability are particularly essential. This transparency extends to the design conditions and the identity of the system designers.

These considerations also highlight the need to always carefully consider the role of AI in the decision-making process (e.g. as illustrated before, in certain situations AI systems should play a supportive role only).

### General Comments

- The development of Artificial Intelligence, automation systems and other emerging technologies bring new challenges in terms of liability and data access and those issues should be carefully addressed and checked whether the current legal framework is adequate. In this regard, the CCBE

is part of the Commission' Expert group on Liability and new technologies which aims is to provide the EU with expertise on the applicability of the Product Liability Directive to new technologies.

- We would also like to bring attention to the HLEG that, from the point of view of legal practitioners, the CCBE suggested to the Council e-Justice Working Party the idea to establish a set of recommendations on the use of AI in the Justice field which has recently been included in the e-Justice Action Plan for the period 2019-2023.
- The CCBE would welcome the possibility of contributing to this discussion on the issues around the use of AI applications and its possible impact from the point of view of Justice and legal practitioners.