A questionnaire to consult European experts on Trustworthy AI

Cristian Baurré, Atia Cortés, Francesca Foffano, Long Pham, Teresa Scantamburlo, Risto Uuk, Florian Zimmermann,

In the road towards Trustworthy AI, a major challenge is to take concrete actions to move from principles to practice and make the ideal of trustworthiness a reality. Not surprisingly, experts take a different stance on what building Trustworthy AI means and propose different strategies to achieve it. Some believe that the development of Trustworthy AI rests, first and foremost, on engineering ethical principles, as, for instance, software toolkits that can help the scrutiny of particular ethical requirements. Other more systematic approaches are concerned with the design of artificial moral agents, such as machine ethics (Anderson and Anderson, 2011). Other experts hold that a purely engineering approach suffers from severe limitations. For example, Arvan (2018) argued that existing methods to programming ethical AI are either too semantically strict, too semantically flexible or overly unpredictable.

While the debate between different conceptualizations of Trustworthy AI goes on in the background, the AI ethics community at large (academia, non-profit organisations, companies, etc.) has provided a multitude of methods and tools which vary in the strategy adopted and the purpose they want to achieve (see Morley et. al (2020) for an extensive review). The abundance of methods, policy options and the recent EU proposal for a regulation (EC, 2021) have added a layer of complexity to the debate about Trustworthy AI introducing further considerations with respect to firms and authorities, among others.
In the context of the AI4EU project, and more specifically the Observatory for Society and AI (OSAI), a group of researchers created a questionnaire examining experts' views on three fundamental trajectories:

1. *General approach to Trustworthy AI*: this set of questions aims to investigate experts' opinion about different approaches to the notion of Trustworthy AI. For example, is this a purely technical concept or are non-technical methods also important? In addition, this part addresses two important components of the European ethics guidelines, i.e. interdisciplinary work and stakeholder participation.
2. *The implementation of Trustworthy AI:* these questions explore the experience and best practices in the field. This includes, for example, the evaluation of feasibility in achieving AI principles (AI HLEG, 2018), the level of confidence with existing methodologies / tools and experts' opinion on the utility and relevance of such tools. Also, there are questions related to the importance of promoting collective discussion and ethical reflection among engineers and computer scientists.
3. *The governance of AI:* this group of questions aims to gain information and opinions about governance mechanisms to achieve Trustworthy AI. This includes, among others, the evaluation of the Proposal for a Regulation on AI and opinions on soft law mechanisms.

The questionnaire addresses a broad range of experts who deal with Trustworthy AI from different angles and fields of study such as Computer Science, Engineering, Philosophy, Law and Political Science. The questionnaire is available online at this web address: https://www.consultationai4eu.eu/

In the recent past, there have been other consultations relating to AI and its social and ethical issues. For example, Muller and Bostrom (2016) examined experts' predictions on the development of high-level machine intelligence and the associated risk for humanity in the coming decades. Grace et al. (2018) asked machine learning experts about their predictions on the progress in AI, with the aim to connect policymakers with the opinion of researchers. In 2020, after the publication of the White Paper on AI, the Commission made available a public consultation addressed to AI practitioners, public and private sectors, SMEs, academia and citizens. The aim was to collect feedback on the upcoming policy options presented in the document. The Ad-Hoc Committee on AI (CAHAI, 2021) launched a multi-stakeholder consultation to identify the key elements of the legal framework based on the Council of Europe's standards on human rights, democracy and the rule of law. The AI-CLAIRE association prepared a survey for the AI community and general audience regarding the proposal for regulation on AI. Both consultations share a main interest in the regulatory process of AI and are addressed to a wide spectrum of participants.

The questionnaire created by this working group targets only AI experts from different fields of domain, aiming to understand their vision on the notion of Trustworthy AI as well as their familiarisation with the existing methods to implement ethical requirements into practice. It is up to date with the latest actions taken by the EU Commission on defining the regulatory framework for AI, but also includes broader aspects related to the culture of AI and ethics in Europe. We hope to raise interest among AI experts, particularly participants to H2020 funded projects, to contribute to this survey with their opinions. We expect to reach a large sample of answers and publish the results before the end of 2021. This survey, along with a citizen consultation created by the Social Awareness working group, will strengthen one of the main objectives of the OSAI, which is to bridge the knowledge gap existing today with AI practitioners and AI users.

References

AI-CLAIRE (2021), *Response to the European Commission's Proposal for AI Regulation and 2021 Coordinated Plan on AI*, https://claire-ai.org/wp-content/uploads/2021/08/CLAIRE-EC-AI-Regulation-Feedback.pdf

AI HLEG: High Level expert Group on AI, Ethics Guidelines for Trustworthy AI (2018), https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Anderson, M. and Anderson, S. (2011), *Machine Ethics*, Cambridge University Press, https://doi.org/10.1017/CBO9780511978036

Arvan, Marcus (forthcoming). Mental time-travel, semantic flexibility, and A.I. ethics. _AI and Society_:1-20.

CAHAI: Ad-Hoc Committee on AI (2021), *Analysis of the Multi-stakeholder Consultation*, https://rm.coe.int/cahai-2021-07-analysis-msc-23-06-21-2749-8656-4611-v-1/1680a2f228

EC: European Commission, Proposal for a 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 (2021), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., Evans, O. (2018), Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts, Journal of Artificial Intelligence. https://doi.org/10.1613/jair.1.11222

Müller, Vincent C. & Bostrom, Nick (2016). *Future progress in artificial intelligence: A survey of expert opinion.* In Vincent Müller (ed.), Fundamental Issues of Artificial Intelligence_. Springer. pp. 553-571.

Morley, J., Floridi, L., Kinsey, L. *et al.* From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Sci Eng Ethics* 26, 2141–2168 (2020). https://doi.org/10.1007/s11948-019-00165-5