SIENNA Submission to the Consultation on the White Paper on Artificial Intelligence

13 June 2020

Introduction

This document provides feedback on the European Commission White Paper on Artificial Intelligence based on the findings and results of SIENNA, a European Horizon 2020-funded project.[1] It accompanies the SIENNA response to the consultation questionnaire. Both submissions reflect only the views of contributors[2] that have prepared this input based on their research in the SIENNA project.

SIENNA (Stakeholder-Informed Ethics for New technologies with high socio-ecoNomic and human rights impAct) is looking into ethical, legal and human rights issues and is developing ethical guidelines for human genomics, human enhancement and AI & robotics. It has received funding under the European Union's H2020 research and innovation programme under grant agreement No 741716.

Recommendations

1. Put fundamental rights and societal values first

The focus of the regulatory framework should not be to build consumers' and businesses' trust in AI, and therefore speed up the uptake of the technology. The objectives and focus of any new EU regulatory framework on AI should be clear and more driven by fundamental rights and societal values. Currently as framed, the vision seems to be one that pushes the uptake of the technology directly or more greatly – with socially, environmentally and economically optimal outcomes and compliance with EU legislation, principles and values as a sideshow. While flexibility will be needed, the intention to limit changes to 'clearly identified problems for which feasible solutions exist' might be very limiting in terms of what is required from the law.

2. Supplement the risk-based approach with an open-ended clause to avoid loopholes

On a general level, the proposed risk-based approach to the new regulatory framework for AI constitutes a reasonable measure aimed at providing a proportionate legislative intervention. While acknowledging that legal certainty requires that there are clear criteria for determining high-risk AI applications, the proposed approach of providing closed lists (of high-risk sectors and of high-risk

[1] SIENNA project: https://www.sienna-project.eu.

[2] Konrad Siemaszko (Helsinki Foundation for Human Rights), Rowena Rodrigues, Anaïs Rességuier (Trilateral Research); Javier Valls Prieto (University of Granada).

purposes of applications irrespective of sectors) may be problematic. To avoid the danger of leaving some high-risk applications under-regulated, this approach could be supplemented with an open-ended clause, for instance, with a requirement to conduct a human rights impact assessment (HRIA) or other relevant impact assessment (e.g., ethical impact assessment as outlined in CWA 17145 Part 2[3]) for other applications, that are also likely to result in medium to high-risk to individuals, communities, and society at large. This is especially important having in mind unintended effects and consequences, that level of risks may evolve over time and that new applications of AI may be quickly developed – and a closed list may fall short and lag behind.

3. Enhance protection of vulnerable groups and individuals

The criteria for assessing the risk of AI applications (be it within the high-risk sectors or outside of them) should acknowledge the needs of and impacts on vulnerable groups or individuals, who might be especially affected by the adverse impacts of AI,[4] such as people with disabilities, patients, children, minorities, migrants[5], elderly, disfavoured or 'excluded' people or social welfare recipients.[6] E.g., AI-based or robotic deception has huge impact on the vulnerable. The vulnerable should not be left behind and AI 'divides' must not be fostered in any way.[7] The White Paper as a whole seems to overlook the issue of adequate protection of vulnerable groups or individuals vis-à-vis adverse impact of AI applications (briefly mentioning "rights of special groups, such as children, older persons and person with disabilities" only in a footnote in the context of risks of remote biometric identification). The concept of vulnerability must also be agile and under review to take into account the impacts of different applications and uses of AI and technological developments.

4. Expand the scope of concerns on the AI impacts

Although it is understandable that not all potential harms can be listed in the White Paper (i.e., that the outline of concerns is non-exhaustive), there are highly important additional concerns missing in the "Problem definition" section. It should also refer to impacts on ability to enjoy fundamental rights and values including social and economic rights (e.g., right of self-determination, right to work, right of everyone to the enjoyment of just and favourable conditions of work, right of everyone to social security, right to education, right to the enjoyment of the highest attainable standard of physical and mental health etc.). Moreover it should not overlook impacts on democratic processes (e.g. on voting

[3] CEN Workshop Agreement 17145, Ethics assessment for research and innovation - Part 2: Ethical impact assessment framework, 2017, https://www.cencenelec.eu/research/CWA/Pages/default.aspx.
[4] Jansen, Philip, et al., 'SIENNA D4.1: State-of-the-art Review: AI and robotics', April 2018, https://www.sienna-project.eu/digitalAssets/787/c_787382-l_1-k_sienna-d4.1-state-of-the-artreview--final-v.04-.pdf.
[5] European Parliament Committee on Legal Affairs, Draft report with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL).
[6] SIENNA and SHERPA, Commentary on the European Parliament Committee on Legal Affairs Draft report with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL).
[7] Rodrigues, Rowena et al., 'SIENNA D4.2: Analysis of the Legal and Human Rights Requirements for AI and Robotics in and Outside the EU', 2019. https://www.sienna-project.eu/digitalAssets/801/c_801912-l_1-k_siennad4.2-legal-analysis-ai-robotics-awaiting-approval.pdf

and elections), issue of disinformation supported by AI, potential loss of access to justice, detrimental effect on rule of law, as well as potential loss of liberty and harms from restricting freedom of movement (in cases of wrongful imprisonment or deportation). As outlined in the White Paper and SIENNA research, environmental impacts must also be duly considered as these are critically important to address.

5. Add the requirement of Human Rights Impact Assessment

Human Rights Impact Assessment should be added as a general requirement for high-risk applications of AI, including the obligation to identify measures to mitigate risks to fundamental rights – as required *inter alia* by the Council of Europe Committee of Ministers recommendation on the human rights impacts of algorithmic systems.[8] This would impose an additional obligation to adopt safeguards or other mitigation measures that are tailored to a specific AI application in question, beyond the six general requirements listed in the White Paper.

6. Set more definitive red lines on what is not permissible

It is important to set certain red lines or ground rules on what applications should not be permitted as clearly violating EU fundamental rights and values. These should include a ban on AI-enabled large-scale scoring of individuals,[9] a ban on AI-based racial profiling systems and a ban on biometric recognition facilitating mass surveillance (understood as a surveillance that is indiscriminate, not targeted against a specific individual[10]). The requirements for remote biometric identification outlined in the current version of the White Paper do not provide sufficient guarantees against abuses – especially vulnerable people and minority groups are at severe risk of misuse (overt and covert) of biometric recognition systems and resulting wrongful decision-making.

7. Construct voluntary labelling scheme for no-high risk AI applications with caution

If properly implemented, voluntary labelling for no-high risk AI applications can be one way to enhance trust and verify that compliance with certain rules (if appropriately aligned with the industrial and societal standards of different contexts). But it must not be seen as *replacement* of responsibility. The schemes must also not become self-serving (i.e., serving the profit interests of the certified). If public administrations are involved in setting up or managing this certification scheme, they would incur scheme design, implementation, monitoring (oversight) and enforcement burdens. Another challenge would be getting organisations to certify. Further its feasibility and sustainability will depend on sustained efforts and support from governments/public sector to incentivise its creation and then

[8] Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, 8 April 2020.
[9] High-Level Expert Group on Artificial Intelligence, *Policy and Investment Recommendations for Trustworthy Artificial Intelligence*, Brussels, 26.09.2019, p. 20, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60343.
[10] EDRi, *Ban Biometric Mass Surveillance. A set of fundamental rights demands for the European Commission and EU Member States,* Brussels, 13.05.2020, https://edri.org/wp-content/uploads/2020/05/Paper-Ban-Biometric-Mass-Surveillance.pdf .

effective use. It also depends on whether certified schemes are able to achieve higher market penetration, whether the schemes have a strong (technical or regulatory) framework and is non-ambiguous; buy-in to the scheme and trust in it. Such voluntary labels must not become a business-manipulated front for hiding risks and harms from non-high risk AI applications and their value varies depending on the nature of the standard, certification or accreditation, underlying criteria, and at whom they are targeted. Risks of forgery of certificates have also to be considered. What is more, voluntary labelling schemes might also create further fragmentation (this depends on the new legal instrument that sets out the voluntary labelling framework).

8. Adopt rigorous prior assessment for high-risk applications

Prior conformity assessment for high-risk applications should be rigorous, not symbolic in compliance and auditing. There should be provisions for appropriate procedures, governance, and legislation (hard law, soft law, sectoral agreements) to support it (there is a need to be 'motivated' about certification). Support of sectoral regulators is also very critical in this context, given their lack of resources Other incentives include sanctions for non-compliance (greater than mere recall of certification). For further information, refer to the results of the EU-funded SATORI project that explored the potential of conformity assessment techniques to support ethics assessment.[11]

9. Avoid blanket promotion of adoption of AI without a necessity and proportionality assessment

The White Paper acknowledges that the use of AI brings both opportunities and risks, but at the same time it states that "*it is essential that public administrations, hospitals, utility and transport services, financial supervisors, and other areas of public interest rapidly begin to deploy products and services that rely on AI in their activities*". Although it may be very beneficial to facilitate and actively support such adoption where it is appropriate, the assessment of this appropriateness should be done on a case by case basis with due regard to the societal context and its sensitivities. It should be taken into account that all of above listed are high-risk sectors. Any deployment of products and services that rely on AI, and especially a rapid deployment, must be accompanied by adequate and sufficient measures to protect the fundamental rights and societal values.

10. Promote ethics by design approach

Ethics by design means incorporating ethical guidelines, recommendations and considerations into design and development processes of artificial intelligence, robotics and related technologies. This methodology fills a gap in current research ethics approaches, which are often too general and abstract for developers to meaningfully understand and apply. 'Ethics by design' methodologies identify how, at different stages in the development process, ethical considerations can be included in development, by finding ways to translate and operationalize ethical guidelines into concrete design practices. An

[11] SATORI, Exploring the potential of conformity assessment techniques to support ethics assessment, Deliverable 7.2, 2017. https://satoriproject.eu/media/D7.2-Exploring-the-potential-of-conformityassessment-techniques-to-support-ethics-assessment.pdf

extensive ethics by design approach for AI has been developed as part of SIENNA and SHERPA[12] projects. [13] Ethics by design approach could be taken into account, among others:

- in the context of "Skills" section of the White Paper, as part of "transforming the assessment list of the ethical guidelines into an indicative «curriculum» for developers of AI that will be made available as a resource for training institutions.";
- as a way of providing more details to mandatory requirements from the section 5 (d) or
- used in the context of the voluntary labelling for non-high risk AI applications (it could be relevant for potential specific set requirements "especially established for the purpose of the voluntary scheme" that is mentioned in the section 5(g).

11. Match funding in AI R&D with funding to study ethical, social and human-rights aspects of AI

Funding in AI R&D should be matched with funding to study ethical, social and human-rights aspects and challenges and carry out full-blown and strong impact assessments of AI (in the short, medium and long term). Further the creation of testing and experimentation sites to support the development and subsequent deployment of novel AI applications should include consideration of ethical and human rights challenges and impacts using appropriate methodologies (e.g., impact assessments). The planned masters programmes in AI should include or integrate modules on human rights, ethics and social impacts of AI. This should come together with a valorisation of the social sciences and ethics of AI, which should be further promoted (including through increase in funding) at various levels, including in social sciences and ethics departments, in engineering schools and in the industry.

12. Support education of the public on AI

The current level of understanding is very low. Without an improvement of this level of understanding of AI on the part of the general EU population, it is difficult to promote trust for the technology. Trust implies the delegation of control and responsibility; in order to ensure this delegation is appropriate and in according to fundamental rights and values in the EU, a certain degree of understanding of the processes at stake is necessary.

13. Replace the term 'citizens' with 'individuals', 'the public' or 'everyone'.

'Citizen' is a highly contested social and political concept, and the use of this language may unnecessarily detract from the value of the proposed framework. Non-citizens in Europe, which includes migrants, refugees and stateless people, are particularly vulnerable to the adverse impacts of artificial intelligence, robotics and related technologies; for example, these technologies are being

[12] Macnish, Kevin, Mark Ryan, 'D3.2 Guidelines for the development and the use of SIS', SHERPA. https://doi.org/10.21253/DMU.11316833.v1. More about SHERPA project, see: https://www.project-sherpa.eu/about/

[13] SIENNA and SHERPA Commentary on the European Parliament Committee on Legal Affairs Draft report with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL).

developed and deployed in border security, crime and terrorism and public service provision. Removing general references to 'citizens' would also bring the White Paper in line with other legislative measures, including the Charter of Fundamental Rights of the European Union (where general provisions apply to 'everyone' and specific citizen rights are explicit) and the GDPR (which does not mention 'citizens'). The proposed European regulatory framework for trustworthy AI to be truly "solid" must be inclusively framed to protect everyone in Europe and as high-risk sectors include asylum, migration, border controls and judiciary, social security and employment services (as per the White Paper), it is critical that the proposed framework protects the human rights of all.

14. Protect whistleblowers working in the AI field

Protection of whistleblowers reporting on the abuses in the process of developing, deploying or using AI applications could significantly contribute to the effective enforcement of the EU regulatory framework for AI. Therefore the White Paper should envisage an appropriate amendment of the scope of the Directive (EU) 2019/1937 on the protection of persons who report breaches of Union law, as proposed in the European Parliament Committee on Legal Affairs Draft report with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL).