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)*

Coordinated by Trilateral Research
based on research from the SHERPA and SIENNA projects.

22 May 2020

This document provides feedback 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) based on the findings and results of the SHERPA and SIENNA EU-funded projects. This commentary reflects the views of contributors[1] that have prepared this input based on their research in the SHERPA and SIENNA projects. It does NOT reflect the views of the European Commission.

SHERPA (Shaping the ethical dimensions of smart information systems (SIS) – a European perspective) analyses how artificial intelligence (AI) and big data analytics impact ethics and human rights. 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. The SIENNA project has received funding under the European Union's H2020 research and innovation programme under grant agreement No 741716. The SHERPA project received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

This document contains: general recommendations; recommendations on a European Agency for Artificial Intelligence, a European certification of ethical compliance, and national supervisory authorities; and specific recommendations to language in the draft *Regulation on ethical principles for the development, deployment and use of artificial intelligence, robotics and related technologies*.

General Recommendations

Recommendation 1: Justify why a new Regulation should be framed in the context of ethical principles.

The context and rationale for framing a Regulation on ethical principles must be clearly explained and justified, especially since requiring compliance with ethical principles in the manner written in the current draft is a novel approach and departure from established legislative practice. A grounded basis for this approach is very important to provide clarity for all stakeholders, reduce concerns related to legitimacy, practical application, and enforceability, and ensure that existing human rights frameworks are not undermined.

[1] Rowena Rodrigues, Nicole Santiago, Anais Resseguier (Trilateral Research); Bernd Stahl (De Montfort University); Konrad Siemaszko (Helsinki Foundation for Human Rights); Stéphanie Laulhé Shaelou (University of Central Lancashire Cyprus).

Recommendation 2: Make protection from harm the primary objective and prioritize the principle of non-maleficence, i.e., an obligation to do no harm.

A Regulation for artificial intelligence should have as a primary aim the protection of people and society from the harmful and adverse impacts of artificial intelligence, robotics and related technologies.

The European Parliament should not be seen or perceived to echo industry 'trust' ambitions. Trust has very different meanings among stakeholders from technical developers to ethicists, and it is therefore an inadequate basis for effective governance. Furthermore, the "robustness, resilience, security, accuracy and error identification, explainability and identifiability" (Recital 11 of the draft Regulation) of these technologies are insufficient to ensure that ethical principles and human rights, including human dignity, justice, and rule of law, are respected and protected.

The potential for misuse and abuse of artificial intelligence, robotics and related technologies must be explicitly acknowledged. Therefore, in addition to the concept of 'good faith', developers, deployers, and users should be held to the ethical principle of non-maleficence. This Regulation is a prime opportunity for the Parliament to take a lead in ensuring that artificial intelligence, robotics and related technologies take into account and are in line with ethical values, and work with society and for society.

Recommendation 3: Explicitly state the applicable ethical principles.
Article 5 (Ethical principles of artificial intelligence, robotics and related technologies) of the draft Regulation does not explicitly outline all the binding ethical principles applicable to developers, deployers and users. While there are many ethical issues referenced throughout the draft Regulation, there is no compiled **clear** list that defines which ethical principles are binding and which are aspirational and voluntary. Without clarity, developers, deployers, and users will not understand their obligations under the Regulation.

Recommendation 4: Define 'High-Risk' AI, robotics and related technologies, taking into account unknown and intended consequences.
A risk-based approach to AI, robotics and related technologies is appropriate and resonates with related approaches, e.g., in data protection. The determination of what is a high-risk AI application should be clear, easily understandable, and applicable for all parties concerned. However, the language in the draft Regulation does not clearly or consistently define 'high-risk' technologies. In Recital 9 of the draft Regulation, a 'high-risk' technology "entails a high risk of breaching the principles of safety, transparency, accountability, non-bias or non-discrimination, social responsibility and gender balance, environmental friendliness and sustainability, privacy and governance." Later, in Article 7 (Risk assessment) of the draft Regulation, 'high-risk' technologies are those that "entail a significant risk of breaching the ethical principles." These two definitions are not consistent.

A clear definition of 'high-risk', which includes reference to specific ethical principles and a definition of 'significant', should be included in Article 4 (Definitions) of the draft Regulation and referenced consistently in the Regulation. This definition should acknowledge that

unknown risks and unintended consequences are a key and significant concern. Relying on self-reported risks related to the intended use of AI, robotics and related technologies by developers, deployers, and users during the assessment process is unlikely to capture realistic concerns and likely hard-hitting impacts.

Recommendation 5: Clarify which provisions of the Regulation apply only to 'High-Risk' technologies.
Article 7(2) (Risk assessment) and Article 14(1) (Supervisory authorities) in the draft Regulation suggest that only those technologies considered to be 'high-risk' will be obligated to comply with the ethical principles set out in the Regulation. In contrast, Article 8(1) (Safety features, transparency and accountability) states that "any" artificial intelligence, robotics and related technologies, including software, algorithms and data used or produced by such technologies, must comply with the ethical principles. This inconsistency is confusing and would likely frustrate effective implementation of the Regulation. Developers, deployers, and users need clarity on which ethical principles are legally-binding within the Regulation on high-risk technologies and which are legally-binding on non-high-risk technologies. Additionally, if exemptions are created, they should be very narrowly construed.

Recommendation 6: Create and/or promote an effective complaint and redress mechanism accessible to stakeholders.
Stakeholders, particularly affected users and members of the general public, need a way to report concerns and gain redress for breaches of the ethical principles outlined in the Regulation. This mechanism could complement other complaint and redress tools, including GDPR administrative complaints and national institutions protecting fundamental/human rights. In creating and/or promoting such a mechanism, consideration should be given to effective investigatory and effective enforcement powers, which may include mitigation strategies, including stopping the development, deployment, or use of certain technologies.

Recommendation 7: Encourage the use of 'ethics by design'.
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 extensive ethics by design approach for AI has been published as part of SHERPA. SIENNA is building on this to present an extended approach for ethics by design that has wider applicability.

Recommendation 8: Promote and fund ethics and human rights education, training, and research on social and ethical impacts.
For all members of the AI ecosystem to be able to understand ethical and human rights concerns, ethical principles and human rights need to be integrated into education and training at all levels for all stakeholders. As a society, we need further understanding and

education on the social impacts of these technologies, especially as many social impacts are not immediately or easily discernible.

While developers, deployers and users will require different bodies of knowledge, educational and training material needs to be created and used, and incentives for their integration developed (e.g., professional accreditation, liability legislation). This should include training on the social sciences of technology (e.g., the fact that data is never 'neutral' or 'raw', but that it reflects particular choices and structural/historic inequalities). Exchanges between the European Agency on Artificial Intelligence, the supervisory authorities, and, where appropriate, with the supervisory authorities of third countries or with international organisations, should be encouraged and facilitated. Adequate funding should be encouraged/facilitated at EU and Member State levels to ethical and human rights education, training and research alongside funding of R&D for such technologies.

Recommendation 9: Establish a trusted position to oversee AI ethics.
Developer, deployer, and user organisations should have an internal position that combines scientific understanding, ethical and human rights awareness to advise on the development and use of AI systems. When there is a concern, this person or team must have necessary independence to speak out to prevent or mitigate risks of harm. The position of a data protection officer (DPO) as stipulated in the General Data Protection Regulation (GDPR) can serve as an example. These ‘AI protection officers’ could be required in high-risk cases and only encouraged in others. Such a role may be aligned and combined with existing positions, e.g., data protection officer, Chief Ethics Officer, Corporate Social Responsibility (CSR) Officer or Business and Human Rights Officer. Qualification standards and education pathways for this position should be clarified.

Recommendation 10: Clarify how conflicts in the requirements of the Regulation and other existing ethical and/or legal frameworks will be addressed.
The Motion and draft Regulation text refer to existing European law and relevant governance standards, but it is not clear how conflicts and discrepancies will be resolved.

Recommendation 11: Replace ‘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 this Regulation. 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 developed and deployed in border security, crime and terrorism and public service provision. Removing general references to ‘citizens’ would also bring this Regulation 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’).

Recommendation 12: Expand the list of vulnerable groups.
When mentioned, the list of vulnerable groups should include the elderly, disfavoured or ‘excluded’ people, inhabitants of poor countries, and social welfare recipients.

Recommendation 13: Remove reference to ‘human-made AI’.
We have already created AI-systems that have built their own AI-systems. Therefore, exclusive concern with ‘human-made AI’ is becoming outdated. To help ensure the Regulation continues to be relevant in the future, reference should only be made to ‘human-centric’ AI.

Recommendation 14: Remove references to 'free will'.
Despite the language in the Explanatory statement that "the concept of free will, an inalienable feature of humanity, does not appear to be in danger at the moment," free will is already very threatened by artificial intelligence, robotics and related technologies. Free will represents "the power or capacity to choose among alternatives or to act in certain situations independently of natural, social, or divine restraints." One could find many examples of how artificial intelligence, robotics and related technologies threaten human free will, and it is misguided to assume differently.

Recommendation 15: Ensure consistency in EU approaches in relation to terminology and scope of application.
There is a concern with the use of different terminologies and scopes of application of proposed new regulatory measures (E.g., 'AI-system' versus 'artificial intelligence, robotics and related technologies'). If/Where intentional, these differences should be clarified to reduce confusion. Consistency of approach should also be ensured between the EU institutions.

Recommendations concerning a European Agency for Artificial Intelligence

SHERPA is currently investigating the feasibility of a new/bespoke regulator for artificial intelligence and big data and its terms of reference; this should inform the work of the Parliament. If created, the following recommendations should be taken into account:

Recommendation 16: Clarify the role of a European Agency for Artificial Intelligence.
To ensure effectiveness of the framework, and to avoid fragmentation, a clear governance structure must be established. Its remit and relationship with sectoral regulators must also be clarified further. In the draft Regulation, it is not clear which entity has ultimate authority for enforcement of the ethical principles. Without a clear identification of responsibility, the resulting confusion and deference will likely undermine any attempts at implementation and enforcement. Consideration should be given to the Agency's level of independence (I.e., it would be preferable for it to be independent from the European Commission and directly report to the European Parliament). Additionally, issues related to duplication of work – e.g., with guidance issued by bodies such as the High-Level Expert Group on Artificial Intelligence (AI HLEG), European Data Protection Supervisor, European Data Protection Board, Court of Justice and the Fundamental Rights Agency – must be addressed.

The responsibility for establishing governance standards and providing professional and administrative guidance should not be delegated to Member States' or national supervisory authorities if EU harmonisation and consistency are desired. Additionally, consideration should be given to whether the standards and guidance will be binding or non-binding. Furthermore, a European Agency should take the lead in developing an EU-wide list of high-risk technologies. If high-risk assessment is left completely to national supervisory authorities, there is a risk of inconsistency, particularly for developers and deployers present throughout the Union.

Recommendation 17: Broaden the responsibilities of the European Agency for Artificial Intelligence.
In addition, and to supplement the responsibilities already articulated in the draft Regulation, a regulator for artificial intelligence should also:

- evaluate the fitness of purpose of measures designed to comply with the Regulation and its ethical principles; and
- create a constructive collaborative environment for EU AI ethics policy and regulation and promote the adoption of a unified message on the ethics of artificial intelligence, robotics and related technologies to the extent possible/required.

Recommendation concerning a European certification of ethical compliance

Recommendation 18: Carefully and specifically frame certification process.
The proposal for the European certification of ethical compliance is fraught with challenges should consider the following points. (i) The difficulties and challenges inherent in ethical certification should be carefully considered - e.g., it's very problematic to claim a product is certified "ethically compliant". This product might have unknown implications at the time of the certification; it could be used/deployed in harmful ways. (ii) If proceeded with, the common criteria and the application process relating to the grant of a European certificate of ethical compliance to developers/deployers or users seeking to certify the positive assessment of compliance should be carefully considered and set out. (iii) If the *European Agency for Artificial Intelligence* develops the criteria and the process, it should not also be 'the certifier' and its role should be very clear (iv) While certification provides a means of demonstrating compliance, building trust and confidence, gaining reputational and financial advantages, these vary depending on the nature of the standard, certification or accreditation, underlying criteria, and at whom they are targeted (v) Certification should be rigorous not symbolic in compliance and auditing. There should be provision for appropriate procedures, governance, and legislation (hard law, soft law, sectoral agreements) to support it (there is a need to be 'motivated' about certification). 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.

Recommendation concerning the role of Member States and 'Supervisory Authorities'

Recommendation 19: Clarify responsibilities vis-a-vis European Agency for AI and other related organisations and agencies.
National supervisory authorities can play a vital role in protecting people and society from harm created by these technologies, as well as promoting their use for societal and environmental good. However, as discussed above, effectiveness of the proposed regulatory framework depends in part on a clear establishment of responsibilities. A European Agency for AI should be responsible for establishing standards, providing professional and administrative guidance, and creating an EU-wide list of high-risk technologies. The national supervisory authorities' can help ensure implementation of the Regulation and EU-level guidance. Otherwise, Member States will continue to develop individualized standards and guidance, resulting in fragmentation. Consideration should be given to national supervisory

authorities' level of independence, and clarification provided on how a 'margin of manoeuvre' (Recital 4 of the draft Regulation) might be defined and applied.

Recommendations on language in the draft Regulation on ethical principles for the development, deployment and use of artificial intelligence, robotics and related technologies

1. Recital 1:
   - Remove the phrase "based on a desire to serve society". While some artificial

     intelligence, robotics and related technologies are "based on a desire to serve society", this is not true in every case and seems to give undue glorification. These technologies can also be developed with profit, unethical, and/or criminal motives.
   - Clarify what is meant by "comprehensive legal framework of ethical principles" (referring exclusively to the Regulation or to others in addition?).

2. Recital 2: Revise the last sentence to: C*onsistent and homogenous application of the rules set out in the Regulation should be ensured throughout the Union* (in line with framing in other EU Regulations).

3. Recital 9: Add principles: **non-maleficence** (avoidance of misuse/harm), **human dignity, justice, rule of law, equality**. The risks to these are high.

4. Recital 14: Revise to state that users should not misuse artificial intelligence, robotics and related technologies.

5. Recital 20: Align this with Recital 19. **Legitimate aims** have often been misused and/or used as a loophole to facilitate the use of problematic technologies – e.g., in security and law enforcement context, also social welfare and public health surveillance. Clarify what 'objective professional requirements' are.

6. Recital 21: Clarify what is meant by "should perform on the basis of sustainable progress".

7. Recital 25:
   - Revise to add language in **bold**: Socially-responsible artificial intelligence, robotics and related technologies, including the software, algorithms and data used or produced by such technologies, can be defined as technologies which both safeguard and promote a number of different aspects of society, most notably **human well-being**, democracy, **environmental**, health and economic prosperity, equality of opportunity, **social cohesion,** workers' and social rights, diverse and independent media and objective and freely available information, allowing for public debate, quality education, cultural and linguistic diversity, **inclusiveness** and gender balance, digital literacy, innovation and creativity. Socially-responsible technologies are also those that are developed, deployed and used having due regard for their ultimate impact on the physical and mental well-being of people.
   - Replace 'ultimate' with 'adverse' (as ultimate means final).

8. Recital 27: Revise to require that **all** EU-funded research projects include ethical reflection and evaluation (using appropriate methodologies such as ethical impact assessment, ethics by design), and not just the ones dealing with 'social wellbeing'.

9. Recital 33: Include reference to the Law Enforcement Directive (Directive (EU) 2016/680 *of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA OJ L 119*).

10. Recital 37: Clarify what is meant by 'multiple participants' (e.g., multiple participants in ecosystem).

11. Recital 41: Specify what is meant by 'social partners' (e.g., social science experts, ethicists, economists).

12. Article 4: Add definitions of 'ethical', 'good faith', 'high-risk', and 'significant risk'. Language should be consistent with definitions and criteria of high-risk set out in relevant existing regulations (e.g. GDPR) and could be drawn from the work of projects such as SHERPA, SIENNA, and SATORI.

13. Article 7: Define and clarify what constitutes a 'significant risk'.

14. Article 8:
    - (1)(a): Reconsider whether a requirement that these technologies are "*developed, deployed and used in a consistent manner so that they do not pursue aims or do not carry out activities other than those for which they have been conceived*" is unrealistic and constitutes a disproportionate restriction of legitimate creative modifications of software to achieve new outcomes.

    - (1)(h): Revise to add language in **bold**: "in accordance with Article 6(3), developed, deployed and used in a manner that makes it possible, in the event of non-compliance with the safety features set out in subparagraphs (a) to (g), for the technologies concerned to be temporarily disabled and to revert to historical functionalities, **or in cases of severe risks to be permanently disabled**."

15. Article 10:
    - (2)(a): Add "while respecting freedom of expression" or "while refraining from illegitimate restriction of freedom of expression online" to prevent unintentionally creating too strong of incentives for blocking legitimate content and overly restricting freedom of expression on line.

    - (2)(d): Revise to include "inclusive and gender-balanced".

16. Article 10(4): Revise to add language in **bold**: "The social effects of the ubiquitous presence of artificial intelligence, robotics and related technologies, including software, algorithms and data used or produced by such technologies, developed, deployed or used in the Union shall be monitored by the national supervisory authorities referred to in Article 14, in order to avoid disruptive effects **on human well-being**, social agency and social relationships, as well as the deterioration of social skills".

17. Article 12: Adopt a stricter position on biometric recognition, e.g., an explicit ban on biometric recognition that could lead to mass surveillance (understood as a surveillance that is indiscriminate, untargeted, i.e., that is not targeted against a specific individual, that would start with a reasonable suspicion against this individual).

18. Article 17: Revise to require reporting more frequently than every three years, given the rapid developments and nature of the relevant technology.

References

Jansen, P., 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-art-review--final-v.04-.pdf

Macnish, Kevin, Mark Ryan, *D3.2 Guidelines for the development and the use of SIS*, SHERPA. https://doi.org/10.21253/DMU.11316833.v1

Rodrigues, R., ‘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_sienna-d4.2-legal-analysis-ai-robotics-awaiting-approval.pdf

Rodrigues, R., A Panagiotopoulos, B Lundgren, S Laulhé Shaelou, A Grant, *Regulatory options for AI and big data,* SHERPA D3.3, December 2019. https://doi.org/10.21253/DMU.11618211.v2

Ryan, Mark; Kevin Macnish, *D1.4 Report on Ethical Tensions and Social Impacts*, SHERPA, 2019. https://doi.org/10.21253/DMU.8397134.v2

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-conformity-assessment-techniques-to-support-ethics-assessment.pdf

CEN Workshop Agreement 17145 (2017). Ethics assessment for research and innovation - Part 1: Ethics committee and Ethics assessment for research and innovation - Part 2: Ethical impact assessment framework. https://www.cencenelec.eu/research/CWA/Pages/default.aspx

SHERPA project. https://www.project-sherpa.eu/about/

SIENNA project. https://www.sienna-project.eu