Shaping the ethical dimensions of smart information systems– a European perspective (SHERPA)

Supplemental Feedback to the European Commission on its

White Paper on Artificial Intelligence: A European approach to excellence and trust

Based on research from the SHERPA project[1]

June 14, 2020

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

[1] With contributions from Aequitas, De Montfort University, F-Secure, Mutual Shoots, Trilateral Research, University of Twente, and UClan Cyprus.

SHERPA (Shaping the ethical dimensions of smart information systems (SIS) – a European perspective) is a project that focuses on ethical and human rights aspects of smart information systems (artificial intelligence and big data analytics). SHERPA aims to ensure that its recommendations and findings help to move the AI ecosystems in the desirable direction and is requesting feedback from stakeholders to help it further shape recommendations and outputs. The final recommendations from SHERPA are due in April 2021.

In addition to its response to questionnaire for the public consultation, the SHERPA consortium offers the following preliminary recommendations to the Commission. These recommendations will be subject to further consultation and refinement. At the end of this document is a list of SHERPA activities that support and promote these recommendations.

Recommendation 1: Make protection from harm the primary objective and prioritize the principle of non-maleficence, i.e., an obligation to do no harm.

The EC White Paper distinguishes between an ecosystem of excellence and an ecosystem of trust. We recommend caution when using the term 'trust'. Trust has very different meanings among stakeholders from technical developers to ethicists, and it is therefore an inadequate basis for effective governance. Instead, the European approach for artificial intelligence should have as a primary aim the protection of people and society from the harmful and adverse impacts of artificial intelligence. Excellent AI is AI that is ethical, protects human rights, and promotes human flourishing. No amount of technical excellence is sufficient on its own if it disregards ethical values and human rights and leads to mistrust. The protection of ethical values and human rights should be understood as a societal priority, and not only a means to create trustworthy AI. The potential for misuse and abuse of artificial intelligence must be explicitly acknowledged, and developers, deployers, and users should be held to the ethical principle of non-maleficence.

Recommendation 2: Future-proof policy by thinking beyond AI.

Some of the concerns discussed in the context of AI are directly linked to the specific characteristics of narrow AI, i.e., technologies like machine learning and deep neural networks. Many concerns, however, are caused by broader socio-technical networks including input or output technologies ranging from sensor-networks to social media. As a case in point, the EC's (2020b) report on safety and liability implications refers to the Internet of Things (IoT) and robotics, as well as narrow AI. In order to ensure that policy remains relevant beyond the current AI discourse, it should be clarified that AI is part of a cluster of emerging digital technologies to which these measures also apply.

Recommendation 3: Strengthen the core of the framework for cooperation of national competent authorities taking into account SHERPA's work on a new regulator.

Relying on national competent authorities and national/sectoral regulators as proposed in the EC White Paper allows building on existing strengths. However, such a network that cuts across disciplinary expertise and national competences will require a strong central node with legal competences and sufficient funding. It should work closely with the European centres of excellence and testing centres being proposed. SHERPA is investigating the feasibility of a new/bespoke regulator for AI and its terms of reference, which should inform the work of the Commission.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Recommendation 4: 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 human rights. This mechanism could complement other complaint and redress tools, including GDPR administrative complaints and national institutions protecting human rights. In creating and/or promoting such a mechanism, consideration should be given to effective investigatory and enforcement powers, which may include mitigation strategies, including stopping the development, deployment, or use of certain technologies.

Recommendation 5: Establish a trusted position to oversee AI ethics.

AI developers and user organisations should have an internal position that combines scientific understanding, and 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 GDPR can serve as an example. These 'AI protection officers' could be required in high-risk cases and encouraged in others. Such a role may be aligned and combined with existing positions, e.g. DPO, Chief Ethics Officer, CSR Officer or Business and Human Rights Officer. Qualification standards and education pathways for this position should be clarified.

Recommendation 6: 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. 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.

Recommendation 7: 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 different stakeholders 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). 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.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Recommendation 8: 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 approach. Non-citizens in Europe, which includes migrants, refugees and stateless people, are particularly vulnerable to the adverse impacts of artificial intelligence; for example, these technologies are being developed and deployed in border security, crime and terrorism prevention, and public service provision. Removing general references to 'citizens' would also bring this AI approach in line with other regulatory frameworks, 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').

SHERPA Contributions to the AI Ethics and Human Rights Ecosystem

SHERPA, itself part of an ecosystem consisting of other projects (SIENNA, PANELFIT) and numerous research groups and projects, is working on the following activities that can support and promote the recommendations listed above:

1. Contribution to knowledge base and curricula

SHERPA has done extensive empirical and conceptual research allowing the consortium to build up capacity to contribute to the definition of a required body of knowledge and model curricula. In addition to factual knowledge, SHERPA has also produced audiovisual material and artistic representations of key aspects of AI and ethics to allow reaching out to a broader audience.

2. Terms of Reference for regulator

SHERPA has undertaken a review of regulatory options and is working on the terms of reference for a central regulatory body that could be at the heart of the framework of AI regulation.

3. Ethics by design

SHERPA (in collaboration with SIENNA) has developed guidelines for AI developers and AI users that are based on the principles of ethics by design. It has provided an initial training session for about 80 Project and Policy Officers of the EC (December 2019). SHERPA can contribute to the design and delivery of further relevant training.

4. Standardisation

SHERPA members are involved in ISO SC 42 and are contributing to the development of ethically-aware standardisation.

SHERPA aims to ensure that its recommendations and findings help to move the AI ecosystems in the desirable direction and is requesting feedback from stakeholders to help it further shape recommendations and outputs.

For more information, please visit: https://www.project-sherpa.eu

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Contribution ID: e95ba38e-2239-438c-86d4-5c39f14e30ba
Date: 12/06/2020 17:28:23

Consultation on the White Paper on Artificial Intelligence - A European Approach

Fields marked with * are mandatory.

Introduction

Artificial intelligence (AI) is a strategic technology that offers many benefits for citizens and the economy. It will change our lives by improving healthcare (e.g. making diagnosis more precise, enabling better prevention of diseases), increasing the efficiency of farming, contributing to climate change mitigation and adaptation, improving the efficiency of production systems through predictive maintenance, increasing the security of Europeans and the protection of workers, and in many other ways that we can only begin to imagine.

At the same time, AI entails a number of potential risks, such as risks to safety, gender-based or other kinds of discrimination, opaque decision-making, or intrusion in our private lives.

The European approach for AI aims to promote Europe's innovation capacity in the area of AI while supporting the development and uptake of ethical and trustworthy AI across the EU. According to this approach, AI should work for people and be a force for good in society.

For Europe to seize fully the opportunities that AI offers, it must develop and reinforce the necessary industrial and technological capacities. As set out in the accompanying European strategy for data, this also requires measures that will enable the EU to become a global hub for data.

The current public consultation comes along with the **White Paper on Artificial Intelligence - A European Approach** aimed to foster a European ecosystem of excellence and trust in AI and a Report on the safety and liability aspects of AI. The White Paper proposes:

- Measures that will streamline research, foster collaboration between Member States and increase investment into AI development and deployment;
- Policy options for a future EU regulatory framework that would determine the types of legal requirements that would apply to relevant actors, with a particular focus on high-risk applications.

This consultation enables all European citizens, Member States and relevant stakeholders (including civil society, industry and academics) to provide their opinion on the White Paper and contribute to a European approach for AI. To this end, the following questionnaire is divided in three sections:

Democratic Republic of the Congo	Lesotho	Saint Helena Ascension and Tristan da Cunha	Zimbabwe
Denmark	Liberia	Saint Kitts and Nevis	
		Saint Lucia	

* Publication privacy settings

The Commission will publish the responses to this public consultation. You can choose whether you would like your details to be made public or to remain anonymous.

- **Anonymous**
  Only your type of respondent, country of origin and contribution will be published. All other personal details (name, organisation name and size, transparency register number) will not be published.
- **Public**
  Your personal details (name, organisation name and size, transparency register number, country of origin) will be published with your contribution.

☑ I agree with the personal data protection provisions

Section 1 - An ecosystem of excellence

To build an ecosystem of excellence that can support the development and uptake of AI across the EU economy, the White Paper proposes a series of actions.

In your opinion, how important are the six actions proposed in section 4 of the White Paper on AI (1-5: 1 is not important at all, 5 is very important)?

	1 - Not important at all	2 - Not important	3 - Neutral	4 - Important	5 - Very important	No opinion
Working with Member states	○	○	○	◉	○	○
Focussing the efforts of the research and innovation community	○	○	○	◉	○	○
Skills	○	○	○	◉	○	○
Focus on SMEs	○	○	○	◉	○	○
Partnership with the private sector	○	○	○	◉	○	○
Promoting the adoption of AI by the public sector	○	○	○	◉	○	○

Are there other actions that should be considered?

500 character(s) maximum

Promote and fund AI that is ethical, protects human rights, and supports human flourishing; work with international organisations and civil society to identity and address issues and concerns related to the development and use of AI; set specific AI and innovation goals for the deployment of AI in the public sector

Revising the Coordinated Plan on AI (Action 1)

The Commission, taking into account the results of the public consultation on the White Paper, will propose to Member States a revision of the Coordinated Plan to be adopted by end 2020.

In your opinion, how important is it in each of these areas to align policies and strengthen coordination as described in section 4.A of the White Paper (1-5: 1 is not important at all, 5 is very important)?

	1 - Not important at all	2 - Not important	3 - Neutral	4 - Important	5 - Very important	No opinion
Strengthen excellence in research	○	○	○	○	◉	○
Establish world-reference testing facilities for AI	○	○	○	○	◉	○
Promote the uptake of AI by business and the public sector	○	○	○	◉	○	○
Increase the financing for start-ups innovating in AI	○	○	○	◉	○	○
Develop skills for AI and adapt existing training programmes	○	○	○	○	◉	○
Build up the European data space	○	○	○	○	◉	○

Are there other areas that that should be considered?

500 character(s) maximum

Increase financing for start-ups innovating in AI; support innovation in established and larger companies that might have better access to relevant data.

A united and strengthened research and innovation community striving for excellence

Joining forces at all levels, from basic research to deployment, will be key to overcome fragmentation and create synergies between the existing networks of excellence.

In your opinion how important are the three actions proposed in sections 4.B, 4.C and 4.E of the White Paper on AI (1-5: 1 is not important at all, 5 is very important)?

	1 - Not important at all	2 - Not important	3 - Neutral	4 - Important	5 - Very important	No opinion
Support the establishment of a lighthouse research centre that is world class and able to attract the best minds	○	○	○	○	◉	○
Network of existing AI research excellence centres	○	○	○	◉	○	○
Set up a public-private partnership for industrial research	○	○	○	◉	○	○

Are there any other actions to strengthen the research and innovation community that should be given a priority?

500 character(s) maximum

Enhance interdisciplinary research and innovation; Ensure a lighthouse research centre and/or network and /or other institutional mechanism is inclusive for all potential new entrants.

Focusing on Small and Medium Enterprises (SMEs)

The Commission will work with Member States to ensure that at least one digital innovation hub per Member State has a high degree of specialisation on AI.

In your opinion, how important are each of these tasks of the specialised Digital Innovation Hubs mentioned in section 4.D of the White Paper in relation to SMEs (1-5: 1 is not important at all, 5 is very important)?

	1 - Not important at all	2 - Not important	3 - Neutral	4 - Important	5 - Very important	No opinion
Help to raise SME's awareness about potential benefits of AI	○	○	○	◉	○	○
Provide access to testing and reference facilities	○	○	○	◉	○	○
Promote knowledge transfer and support the development of AI expertise for SMEs	○	○	○	◉	○	○
Support partnerships between SMEs, larger enterprises and academia around AI projects	○	○	○	○	◉	○
Provide information about equity financing for AI startups	○	○	○	◉	○	○

Are there any other tasks that you consider important for specialised Digital Innovations Hubs?

500 character(s) maximum

Create certification around AI educational standards & training; help SMEs identify & address related ethical, legal & societal issues through funding, training activities, & connecting to other resources & best practices; promote public procurement of AI applications that are ethical, protect fundamental rights, & support human flourishing; monitor ISO & CEN activities with regard to AI & help formulate interventions where possible in the standardisation process

Section 2 - An ecosystem of trust

Chapter 5 of the White Paper sets out options for a regulatory framework for AI.

In your opinion, how important are the following concerns about AI (1-5: 1 is not important at all, 5 is very important)?

	1 - Not important at all	2 - Not important	3 - Neutral	4 - Important	5 - Very important	No opinion

AI may endanger safety	○	○	○	○	◉	○
AI may breach fundamental rights (such as human dignity, privacy, data protection, freedom of expression, workers' rights etc.)	○	○	○	○	◉	○
The use of AI may lead to discriminatory outcomes	○	○	○	○	◉	○
AI may take actions for which the rationale cannot be explained	○	○	○	◉	○	○
AI may make it more difficult for persons having suffered harm to obtain compensation	○	○	○	◉	○	○
AI is not always accurate	○	○	○	◉	○	○

Do you have any other concerns about AI that are not mentioned above? Please specify:

500 character(s) maximum

Impacts economic & power distribution (e.g. few powerful corporations develop technologies & influence political processes; unequal distribution of resources & wealth); increased risks to vulnerable groups; affects crisis response (e.g. pandemic misinformation); increases surveillance & mass manipulation; undermines rule of law; affects human relations and decision-making; changes the delivery & accessibility of public services; environmental impacts

Do you think that the concerns expressed above can be addressed by applicable EU legislation? If not, do you think that there should be specific new rules for AI systems?

- ○ Current legislation is fully sufficient
- ○ Current legislation may have some gaps
- ○ There is a need for a new legislation
- ◉ Other
- ○ No opinion

Other, please specify

500 character(s) maximum

'Current legislation may have some gaps' and ‘There is a need for a new legislation.’ A comprehensive gaps analysis is needed to identify & prioritise required regulatory reform. However, this is not an excuse for governments to delay actions, especially where AI is having negative effects. One such immediate action should/could be implementing a ban/moratorium on the use of lethal autonomous weapons systems.

If you think that new rules are necessary for AI system, do you agree that the introduction of new compulsory requirements should be limited to high-risk applications (where the possible harm caused by the AI system is particularly high)?

- ○ Yes
- ◉ No
- ○ Other
- ○ No opinion

Additional Comments

500 character(s) maximum

Risk assessments could be carried out for all AI; applications that are not high-risk should have compulsory requirements to be ethical and protect fundamental rights, though the requirements do not need to be as burdensome as for high-risk applications. There are also concerns that the high-risk determination in the White Paper is unlikely to capture realistic concerns about unforeseen/unintended impacts.

If you wish, please indicate the AI application or use that is most concerning ("high-risk") from your perspective:

500 character(s) maximum

AI use in warfare; AI use by law enforcement authorities; AI use in judicial decision-making systems; AI powered/AI lie detectors (e.g. Automated Virtual Agent for Truth Assessments in Real-Time); AI decision-making with no opportunity for human intervention; AI applications based on techno-solutionism (aka using AI to 'solve' the wrong problems – e.g. discouraging use of crumbling public transport systems instead of fixing them).

In your opinion, how important are the following mandatory requirements of a possible future regulatory framework for AI (as section 5.D of the White Paper) (1-5: 1 is not important at all, 5 is very important)?

	1 - Not important at all	2 - Not important	3 - Neutral	4 - Important	5 - Very important	No opinion
The quality of training data sets	○	○	○	○	◉	○
The keeping of records and data	○	○	○	○	◉	○
Information on the purpose and the nature of AI systems	○	○	○	○	◉	○
Robustness and accuracy of AI systems	○	○	○	○	◉	○
Human oversight	○	○	○	○	◉	○
Clear liability and safety rules	○	○	○	○	◉	○

In addition to the existing EU legislation, in particular the data protection framework, including the General Data Protection Regulation and the Law Enforcement Directive, or, where relevant, the new possibly mandatory requirements foreseen above (see question above), do you think that the use of remote biometric identification systems (e.g. face recognition) and other technologies which may be used in public spaces need to be subject to further EU-level guidelines or regulation:

- ○ No further guidelines or regulations are needed
- ○ Biometric identification systems should be allowed in publicly accessible spaces only in certain cases or if certain conditions are fulfilled (please specify)
- ○ Other special requirements in addition to those mentioned in the question above should be imposed (please specify)
- ○ Use of Biometric identification systems in publicly accessible spaces, by way of exception to the current general prohibition, should not take place until a specific guideline or legislation at EU level is in place.
- ○ Biometric identification systems should never be allowed in publicly accessible spaces
- ◉ No opinion

Please specify your answer:

Do you believe that a voluntary labelling system (Section 5.G of the White Paper) would be useful for AI systems that are not considered high-risk in addition to existing legislation?

- ◉ Very much
- ○ Much
- ○ Rather not
- ○ Not at all
- ○ No opinion

Do you have any further suggestion on a voluntary labelling system?

500 character(s) maximum

But, it must be robust enough that it is meaningful, but not too onerous for SMEs without the resources of large companies; specific goals must be clearly identified & communicated; the common criteria & application process should be carefully considered & articulated; labelling should be rigorous, not symbolic in compliance & auditing; attention should paid to the consistency & interoperability w/ other standardization /certification regimes to avoid a proliferation of inconsistent activities.

What is the best way to ensure that AI is trustworthy, secure and in respect of European values and rules?

- ☐ Compliance of high-risk applications with the identified requirements should be self-assessed ex-ante (prior to putting the system on the market)
- ☐ Compliance of high-risk applications should be assessed ex-ante by means of an external conformity assessment procedure

- [ ] Ex-post market surveillance after the AI-enabled high-risk product or service has been put on the market and, where needed, enforcement by relevant competent authorities
- [x] A combination of ex-ante compliance and ex-post enforcement mechanisms
- [x] Other enforcement system
- [ ] No opinion

Please specify any other enforcement system:

500 character(s) maximum

Should include those of self-assessment ex-ante and assessment ex-ante by means of an external conformity assessment procedure. One potential ex-ante enforcement could take place at the time of patent registration (an individual or business filing a patent must affirm their innovation does not in any foreseeable way violate fundamental rights and freedoms).

Do you have any further suggestion on the assessment of compliance?

500 character(s) maximum

At the int'l level, SHERPA research identified as most promising: Binding Framework Convention, CEPEJ European Ethical Charter, & Legislative Framework for independent & effective oversight. At the EU-level: general fund for smart robots & the Common Union registration of robots; algorithmic impact assessments under GDPR; & voluntary/mandatory certification of ADS. At the nat'l level: redress by design & proposed 'specific' legislation. For more, see SHERPA T3.3 on Regulatory Measures.

Section 3 – Safety and liability implications of AI, IoT and robotics

The overall objective of the safety and liability legal frameworks is to ensure that all products and services, including those integrating emerging digital technologies, operate safely, reliably and consistently and that damage having occurred is remedied efficiently.

The current product safety legislation already supports an extended concept of safety protecting against all kind of risks arising from the product according to its use. However, which particular risks stemming from the use of artificial intelligence do you think should be further spelled out to provide more legal certainty?

- [x] Cyber risks
- [x] Personal security risks
- [x] Risks related to the loss of connectivity
- [x] Mental health risks

In your opinion, are there any further risks to be expanded on to provide more legal certainty?

500 character(s) maximum

Discrimination and bias; lack of accessibility

Do you think that the safety legislative framework should consider new risk assessment procedures for products subject to important changes during their lifetime?

- (●) Yes
- ( ) No
- ( ) No opinion

Do you have any further considerations regarding risk assessment procedures?

500 character(s) maximum

Do you think that the current EU legislative framework for liability (Product Liability Directive) should be amended to better cover the risks engendered by certain AI applications?

- (●) Yes
- ( ) No
- ( ) No opinion

Do you have any further considerations regarding the question above?

500 character(s) maximum

Modifications should not apply only to AI applications narrowly, but also e.g. software, IoT and other services in general.

Do you think that the current national liability rules should be adapted for the operation of AI to better ensure proper compensation for damage and a fair allocation of liability?

- (●) Yes, for all AI applications
- ( ) Yes, for specific AI applications
- ( ) No
- ( ) No opinion

Do you have any further considerations regarding the question above?

500 character(s) maximum

Thank you for your contribution to this questionnaire. In case you want to share further ideas on these topics, you can upload a document below.

You can upload a document here:

The maximum file size is 1 MB
Only files of the type pdf,txt,doc,docx,odt,rtf are allowed

Contact

CNECT-AI-CONSULT@ec.europa.eu