Statement of Support (SoS) for funding bodies

Introduction

The SoS is a valuable tool that can support funding bodies in evaluating AI-related proposals. Each funding body's role when financing AI-based projects is essential to ensure that such developments are in the human interest, respect fundamental rights, and do not cause environmental damage. The advances of AI-based systems are too fast to be captured and contained solely by legal regulation; on the other hand, it is undesirable for legislation to operate as an unnecessary brake on an advance that could mean remarkable improvements for humanity. In this sense, the decision of which projects to fund should be guided by their adherence to globally accepted ethical principles on AI and, of course, their compliance with current legal regulations. On top of evaluating the technical strengths of the applicants' proposals, reviewing their Trustworthy AI Statements is critical. This document will help you to do so by providing guidance on which areas should be considered when assessing those statements.

Instructions

This document is a **Statement of Support (SoS)** for you to use while evaluating **Trustworthy AI Statements** from applicants seeking funding to develop AI-based systems. The SoS was developed in conjunction with a self-assessment tool for applicants called **Stop-and-Think**, which sets out the key areas they should review to ensure compliance with ethical and legal standards[1]. As such, the SoS reflects the same areas of assessment, and we recommend that, where possible, applicants be offered Stop-and-Think to make the application of the SoS more efficient. As the SoS encompasses a selection of mandatory and complimentary requirements and principles, it is advisable to be aware of this difference as it is clarified in each step. At the same time, because a call for application may include different kinds of AI systems, for example, high-risk or limited-risk, this tool is better understood as an holistic approach to evaluation of applications. However, this tool could also be adapted and converted into a scoring tool by funding bodies if considered more appropriate.

The SoS has two uses. First, it can be used as *a guide on key aspects of deciding whether each project under consideration meets ethical and legal requirements*. Second, the *SoS allows projects to be compared to decide which are more appropriate according to the ethical and legal parameters evaluated*.

[1] Listed in the Annex of this tool.

Key Areas to Consider when assessing Trustworthy AI Statements

Step 1: Did the applicant correctly classify the AI system proposed following the AI Act Risk Classification?

First, a brief summary of the AI Act classification. The EU AI Act classifies AI systems into four risk categories: Unacceptable Risk, High Risk, Limited Risk, and Minimal Risk.

- **Unacceptable Risk**: AI systems that deploy harmful manipulative "subliminal techniques"; AI systems that exploit specific vulnerable groups (physical or mental disability); AI systems used by public authorities or on their behalf, for social scoring purposes, "Real-time" remote biometric identification systems in publicly accessible spaces for law enforcement purposes, except in a limited number of cases.
- **High Risk**: AI systems that adversely impact people's safety or fundamental rights. The AI Act differentiates between two categories of high-risk systems. Systems used as a safety component of a product falling under EU health and safety harmonization legislation; systems deployed in eight specific areas detailed in Annex III.
- **Limited Risk**: AI systems that interact with humans (e.g., chatbots), emotion recognition systems, biometric categorization systems, and AI systems that generate or manipulate image, audio, or video content (e.g., deepfakes) would be subject to a limited set of transparency obligations.
- **Minimal Risk**: these systems could be developed and used without conforming to any additional requirements.

If the applicant determined correctly the risk level, the first step in the analysis results in a positive result. If the applicant failed to do so, you should pay extra attention to the rest of the steps. In addition, it is key to note that prohibited practices should not be funded.

Step 2: Application for developing a High-Risk AI system

If the AI system is classified as high-risk, by you or the applicant, ensure it complies, among others, with the following requirements:

- **Risk Management System**: Implement a risk management system to identify, assess, and mitigate risks.
- **Data Governance**: Ensure the quality and integrity of the data used. This includes proper data collection, annotation, and handling procedures.

- **Technical Documentation**: Maintain comprehensive technical documentation detailing the system's purpose, design, development, testing, and deployment.
- **Record Keeping**: Create a system that allows automatic recording of events (logs) over the lifetime of the system.
- **Transparency and Information Provision**: Provide clear information to users about the system's capabilities and limitations.
- **Human Oversight**: Design mechanisms that allow human oversight and intervention when necessary.
- **Robustness, Accuracy, and Security**: Ensure your system is resilient, accurate, and secure against potential threats.

While it is not mandatory, implementing the safeguards required for high-risk AI systems in non-high-risk AI systems can be considered best practice. Therefore, if the applicant of a non-high-risk AI system will implement these safeguards, his or her application should be granted extra weight when compared to others.

Step 3: Ethical Considerations

Adhering to ethical principles is critical to complying with guidelines, the AI Act, ALTAI, and other instruments, on which this tool is based. Ensure the Trustworthy AI Statement discusses potential ethical challenges such as biases, misuses, unintended harms, impact on equality, and proportionality between the proposed system and the intended goals. **Remember, ethical considerations go beyond what is legally mandatory. Something can be legal but unethical or illegal but ethical. In the Trustworthy AI Statement, the applicant should explain how and why their proposal is ethically aligned.** The following fundamental principles should be present and developed in the project, and the applicant should state them in their AI Trustworthy Statement:

- **Human agency and oversight:** AI systems should empower human beings and foster their fundamental rights and should be subject to proper oversight mechanisms
- **Technical Robustness and safety:** To avoid unintentional harm, AI systems should be resilient, secure, accurate, reliable and reproducible.
- **Privacy and data governance**: Comply with GDPR and other relevant privacy regulations. Ensure the AI system does not infringe on individuals' privacy rights.
- **Transparency**: the data, system and AI business models should be transparent and individuals need to know they are interacting with an AI

system. The decisions taken by an AI system should be explained and easily understandable for the individual concerned.

- **Non-discrimination and fairness**: Design your AI system to avoid bias and discrimination. Implement measures to detect and mitigate any potential bias in data and algorithms.
- **Societal and environmental well-being**: AI systems should benefit all human beings, including future generations. It must hence be ensured that they are sustainable and environmentally friendly.
- **Accountability**: Establish clear responsibility for the AI system's decisions and actions. Ensure processes are in place for redress and remedy in case of harm or misuse.

Step 4: Transparency and User Awareness

Transparency and user awareness are key to trustworthy AI. The statement you are evaluating should be robust in this regard, ***regardless of the risk classification****.* Revise the Trustworthy AI Statement to be sure that it complies with the standards put forward by the AI Act:

- **Clear Communication**: Inform individuals when they are interacting with an AI system. Provide understandable information about how the AI system makes decisions. For instance, provide individuals with information about
  - When AI technologies are being used;
  - The capabilities and limitations of a given model;
  - The data on which the model was trained;
  - The data used to generate outputs;
  - Whether data is retained (and if so, what and for how long);
  - Avenues to remediate or appeal outputs produced by the model; and
  - Whether user choices can influence system performance.
- **Documentation for Users**: Offer comprehensive documentation and user guides that explain the AI system's functionality, limitations, and correct usage.

Step 5: Sustainability and Societal Impact

Finally, review how the Trustworthy AI Statement refers to proposed project's sustainability and societal impact. In addition, **if you are assessing similar proposals, take this dimension into special consideration to decide which proposal is the best.** For example, you should consider if some of the following environmental challenges and discrimination problems are analyzed in the statement:

- **Environmental Impact**: Aims for energy-efficient algorithms and sustainable practices.

- **Social Impact**: Evaluates the broader societal implications of the AI system. Ensures it contributes positively to society and does not reinforce existing inequalities or create new ones.

Final Consideration: Trustworthy AI Statement Checklist

Before making a decision, ensure you have addressed the following:

1. **Risk Assessment**: Did the applicant correctly determine the risk level? If it is a High-Risk level system, is it proportionate to the goals it has?
2. **Ethical Considerations**: Has the Trustworthy AI Statement went through all the ethical considerations and explained how the applicant will meet them?
3. **Transparency Measures**: If the project is funded, is the applicant ready to provide users with clear information and documentation?
4. **Impact Assessment**: If the project is funded, is the applicant ready to evaluate and mitigate the environmental and societal impacts of your AI system?

Annex

- EU: AI Act and EC ethical guidelines
- AI Impact Assessment. A tool to set up responsible AI projects, Ministry of Infrastructure and Water Management
- Framework Convention on AI
- USA: Blueprint for AI Bill of Rights
- AU AI ethical principles
- CA Responsible use of AI
- IEEE ethically aligned design
- Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems: A Proposal
- Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment
- EU model contractual AI clauses to pilot in procurements of AI
- IEEE CertifAIEd™ – Ontological - Specification for Ethical Algorithmic Bias
- Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems: A Proposal

- CAN/CIOSC 101:2019 - Ethical design and use of automated decision systems: AI Act draft and EC ethical guidelines
- Center for Inclusive Change, Essential Considerations in AI Contracting
- WEF, Guidelines for AI procurement
- WEF, AI Procurement in a Box: AI Government Procurement Guidelines
- Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems (Canada)
- ISO/IEC 42001:2023(en) Information technology — Artificial intelligence — Management system