Stop-and-Think Self-Assessment tool for applicants

Introduction

The following self-assessment tool will help you complete your Trustworthy AI Statement thoroughly and efficiently. Your role as a potential recipient of funds to develop AI-based systems is critical in respecting globally agreed ethical principles. Adverse effects on human rights and the IA environment are a reality. One way to avoid them is to evaluate the project and present a thorough analysis before implementation. You must understand that the Trustworthy AI Statement does not simply state that you will comply with mandatory legal regulations but implies a solid commitment to the ethical development of AI-based systems. This instrument particularly considers the AI Act, the ALTAI, and tools developed by different governments from EU Member States to evaluate the impact of AI systems[1]. Compliance with ethical and legal standards may require the assistance of experts in these areas. Consider consulting with such experts if the system you intend to develop involves human rights impacts.

---

[1] Listed in the Annex of this tool.

Instructions

The **Stop-and-Think** is a guide to be used during the ethical reflection on the proposed AI-based system development and during the drafting of the **Trustworthy AI Statement**. The Stop-and-Think tool will guide you step by step to elaborate on how you comply with the critical ethical and legal areas and not simply state that you will abide by them. Please note that this tool does not preclude you from adding further specifications depending on the type of project for which you are applying for funding.

Key Areas to Consider when preparing your Trustworthy AI Application Statement

Stage 1: Understanding AI Act Risk Classification

Stop: Before preparing your statement, ensure all your group members understand the AI Act Risk Classification.

Think: Read carefully the summary of the AI Act classification and decide according to it.

1. Risk Classification

- o **Understand Risk Levels**: The EU AI Act classifies AI systems into four risk categories: Unacceptable Risk, High Risk, Limited Risk, and Minimal Risk.
- o **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.
- o **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.
- o **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.
- o **Minimal Risk**: these systems could be developed and used without conforming to any additional requirements.

Hint: use the AI Checker https://artificialintelligenceact.eu/assessment/eu-ai-act-compliance-checker/

Stage 2: Identify Your System's Risk Level

Stop: Which classifications mentioned above does your AI system project fall under?

Think: Determine which category your AI system falls into. Remember that high-risk systems include those used in critical infrastructure, education, employment, essential public services, law enforcement, and migration, among others. **In addition, it is key to note that prohibited practices will not be funded.**

Stage 3: Application that includes a High-Risk AI system [for high-risk AI Systems]

Stop: Revise the High-Risk AI System Requirements of the AI Act.

Think: If your AI system is classified as high-risk, 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.

Hint: 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.

Stage 4: Ethical Considerations

Stop: Adhering to ethical principles is critical to complying with guidelines, the AI Act, ALTAI, and guidelines developed by governments and other organisms, on which this instrument is based. Ensure your team 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 your Trustworthy AI Statement, you should explain how and why your proposal is ethically aligned**.

Think: Revise the following fundamental principles when developing your project and 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 systems 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.

Stage 5: Transparency and User Awareness

Stop: Before fully developing the AI system, ensure transparency and user awareness are considered and that AI Act standards are followed, ***regardless of the risk classification.***

Think: Does the project include the following critical points?

- **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.

Stage 6: Sustainability and Societal Impact

Stop: Before delving deeper into the technical aspects of AI system development, stop and reflect on the proposed project's sustainability and societal impact.

Think: Consider the following environmental challenges and discrimination problems that may arise from your project.

- **Environmental Impact**: Consider the environmental impact of developing and deploying your AI system. Aim for energy-efficient algorithms and sustainable practices.
- **Social Impact**: Evaluate the broader societal implications of your AI system. Ensure it contributes positively to society and does not reinforce existing inequalities or create new ones.

Final Consideration: Trustworthy AI Statement Checklist

Before submitting your funding application, ensure you have addressed the following:

1. **Risk Assessment**: Have you classified your AI system's risk level?
2. **Ethical Considerations**: Have you implemented measures to ensure fairness, privacy, and accountability?
3. **Transparency Measures**: If your project is funded, are you ready to provide users with clear information and documentation?
4. **Impact Assessment**: If your project is funded, are you 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