HOW 'TRUSTWORTHY' IS ARTIFICIAL INTELLIGENCE?

What does 'trustworthy' AI mean? When can you use AI 'in a responsible way'?

To help you with these questions, there are certain guidelines for the development and use of AI. These guidelines focus on the ethical and social issues and questions related to AI.

The Knowledge Centre Data & Society summarises the 7 requirements for AI, written by the High-Level Expert Group on AI (AI HLEG).

For each requirement, a number of questions are formulated that enable you to reflect on the trustworthiness of your AI system.

If you require a more detailed approach, the Knowledge Centre Data & Society recently developed the AI Blindspots cards. By using this tool, you proactively reflect on the (ethical and social) decisions and actions you want to take at the start of your project or during the development of your AI system.

Disclaimer: the 7 requirements overlap, we have therefore reduced redundant questions and chose the most pressing questions to reduce complexity.

Knowledge Centre Data & Society (2020). How ethical is AI? brAInfood from the Knowledge Centre Data & Society. Brussels: Knowledge Centre Data & Society

We would like to thank the CLAIRE Research Network for their feedback on this brAInfood.

HUMAN AGENCY & OVERSIGHT

- Is the interaction between your AI system and a human meaningful and relevant?
- Who takes the final decision, the machine or the human? If it is the machine, then there is no human oversight.
- Is your AI system autonomous or self-learning? If so, are there mechanisms of oversight?

TRANSPARENCY

- Can you show how the algorithm is designed and built, and how decisions are being made to your team but also anyone in contact with the system? Make sure that everyone understands that an AI component is part of the system.

ACCOUNTABILITY

- Do you have an overview of all the decisions and trade-offs you have made to create your system?
- Did you identify negative consequences for all parties involved?
- Were you able to reduce negative consequences and are these measures documented?

TECHNICAL ROBUSTNESS & SAFETY

- Which measures will you take if your AI system is attacked or behaves differently than expected, or when it is used for another (unwanted) purpose?
- Is there a chance your AI system will make inaccurate predictions?
- What measures are in place to address inaccuracy?

DIVERSITY, NON-DISCRIMINATION & FAIRNESS

- How will you avoid unfair bias in your AI system? Is it possible for others to indicate bias or discrimination?
- Did you consider how your AI innovation changes access to your service for marginalised groups? If so, do you provide an alternative service for the excluded?
- Can your stakeholders participate in the development and use of your AI system?

PRIVACY & DATA GOVERNANCE

- Does your AI system use personal data? If so, are you aware of the implications and requirements when using personal data? Can you prove that you comply with data protection regulation?
- Do you have oversight mechanisms in place for the collection, storage, processing and use of data?
- Who can access users' data? Do you have data access protocol measures?

SOCIETAL & ENVIRONMENTAL WELL-BEING

- Can you measure the environmental impact of your AI system's life cycle? Do you know how to reduce it?
- Are (end) users aware of the limitations of interacting socially with your AI system and the societal impact of your AI system?
- Could other groups or individuals be indirectly affected by your AI system, beyond your target audience?

brAInfood of the Knowledge Centre Data & Society

CLAIRE