

The seven principles of AI systems

What are they

- **Human agency and oversight:** AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights.
- **Technical Robustness and safety:** AI systems need to be resilient and secure.
- **Privacy and data governance:** Full respect for privacy and data protection, adequate data governance mechanisms must also be ensured.
- **Transparency:** The data, system and AI business models should be transparent.
- **Diversity, non-discrimination and fairness:** Unfair bias must be avoided.
- **Societal and environmental well-being:** AI systems should benefit all human beings, including future generations.
- **Accountability:** Mechanisms must be in place to ensure responsibility and accountability for AI systems and their outcomes.

Human agency and oversight

With the jupyter notebook the people can oversee what goes inside of the model and what's being chosen to include in the model.

Technical Robustness and safety

In this part I will look at how it will be send over to the platform and how it's received. Furthermore I want to obfuscate a part of the dataset so it becomes a bit safer when its possible to get the dataset inside of the model. But in this prototype its not that big of a concern.

Privacy and data governance

Not everything needs to be plain text and needs to be obfuscated, and if possible I need to send it so it's not plain text when send over.

Transparency

There is a lot of comments inside of the code between the actual bot and the model part which makes it clear how the bot works for the most part. What's inside what is being used to decide what will be used.

Diversity, non-discrimination and fairness

I try to include as much as possible of each country without much caring about age and sex. I try to make it so that the model has an even spread of each country.

Societal and environmental well-being

This is trickier to determine because on what I have chosen for dataset. But it can be updated per year to include the new year.

Accountability

I can't really show it but with a message of "send a mail to xxxx@xxx.xx if this you have concerns over how the data is send/received/processed" or something along these lines.