

1. Introduction

Artificial Intelligence has become significantly more powerful in recent years. As technology improves, it is also becoming embedded into more and more systems that we use every day. This leaves us exposed to the decisions made by these systems. Are we able to trust those decisions? Are those decisions right for us?

As people become more exposed to the decisions of AI, they also become more exposed to its impacts. It is right then to focus on the ethical questions surrounding its development and use.

In this tutorial, we look at questions surrounding Trust, Governance, and its environmental impact.

2. Trust

AI systems are used in decision-making systems, recommending content, or filtering job applications. They have been applied to reading medical scans using image recognition technology and, in the future, may aid medical diagnostics.

In order for people to believe in and benefit from those systems, they need to trust them. The five pillars of trust allow us to outline the properties an AI system should exhibit to be considered trustworthy:

The five pillars of trust are:

- Fairness – emphasizes the importance of a lack of bias
- Explainability – What data was used? What methods were used? What is the provenance of the model and data used
- Robustness – nobody can hack the model to benefit one person over another
- Transparency – fact-sheet meta-data should exist, explaining how the model makes decisions, what data was used, and what known biases exist
- Data privacy – AI systems should maintain privacy and regulatory compliance at all stages, from data gathering and transmission through training and output

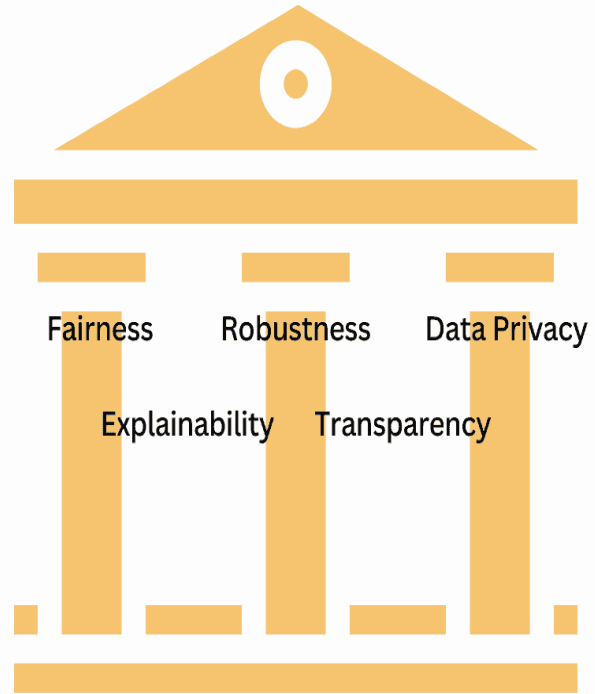
These pillars are required in order for AI systems to become ethically integrated into our lives.

2.1. Fairness

Fairness implies a lack of bias in the model. People need to believe the model acts fairly and that one group of people is not advantaged or disadvantaged compared to another group. Fair technology should not have hidden biases or exacerbate existing biases.

Fairness can be attained through tasks such as data analysis. Ensuring the dataset appropriately covers all cases. For example, ensuring that images sampled for a facial recognition dataset are not predominantly only white males. This was identified as an issue by Joy Buolamwini.

The Pillars of Trust



The five pillars of trust