

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

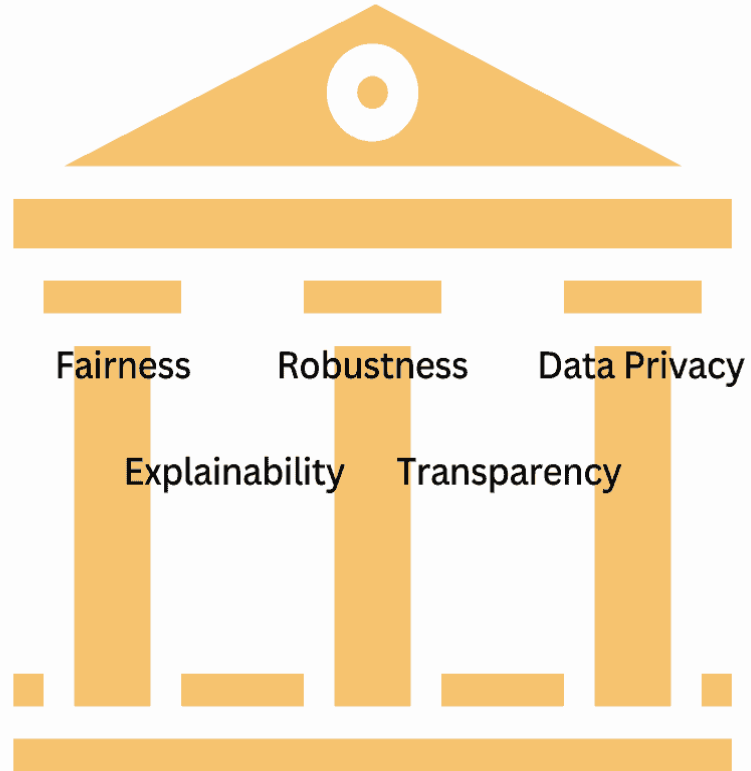
These pillars are required in order for AI systems to become ethically integrated into our lives.

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

2.4. Transparency

Transparency promotes openness, understanding, and accountability in AI systems. It enables scrutiny, evaluation, and ethical decision-making. Transparent AI systems allow for external audits and assessments of fairness, privacy, and other ethical dimensions.

When users and stakeholders have access to information about how AI systems work, including their underlying algorithms, data sources, and processing methods, it helps demystify the technology. Increased transparency allows users to understand better the system's capabilities, limitations, and potential biases. This leads to enhanced trust.

2.5. Data Privacy

Data privacy is important because it is concerned with the responsible use of personal and sensitive information. When individuals have assurance that their personal information is handled securely and that their privacy is respected, they are more likely to trust AI technologies. Data privacy recognizes and respects individuals' rights to privacy and personal autonomy.

Data privacy is closely tied to ethical considerations, particularly in relation to fairness and non-discrimination. Protecting privacy helps prevent the misuse of personal data for discriminatory purposes, profiling, or the reinforcement of biases. Robust privacy measures, such as encryption, access controls, and secure data storage, can reduce the risk of data breaches and ensure fairness. As AI systems often rely on large amounts of personal data, and data breaches can have severe consequences.

3. Governance

We have seen how trust can be built through the five pillars. However, these principles, which outline our ethics, need to be applied. This is where governance is important in ensuring ethical AI development. **Good governance provides a framework for responsible and accountable AI practices, fostering trust, transparency, and ethical decision-making.**

It achieves this through developing frameworks that help establish clear lines of accountability and responsibility for AI development and use. These frameworks help identify and navigate ethical considerations. They help define the legal and regulatory landscape, enable compliance with existing regulations, and inform the creation of new ones. **Good governance establishes data protection, privacy, fairness, transparency, and accountability standards.**

Effective governance ensures that an organization achieves its' responsibilities outlined under the pillars of trust. Governance ensures Fairness and Robustness through policies that mitigate potential risks, such as biases, discrimination, privacy breaches, or safety concerns throughout the entire AI life-cycle.

Governance is a socio-technological challenge that mixes people, processes, and tools. Frameworks for governance define the process, but stakeholders must also actively engage in the process, and appropriate tools must be made available where necessary.

4. Stakeholders and Impact

As AI systems become evermore integrated into various processes, we all become stakeholders in the problem. It matters how AI is used and how it benefits each of us. While there are many potential benefits, there are potential pitfalls, and it is important to consider who is impacted, to what degree, and where the impact falls.

We introduce two major areas where the stakeholders of AI technology go beyond who we might traditionally think of and show how wide-ranging an impact this technology can have.

4.1. Environmental Impact

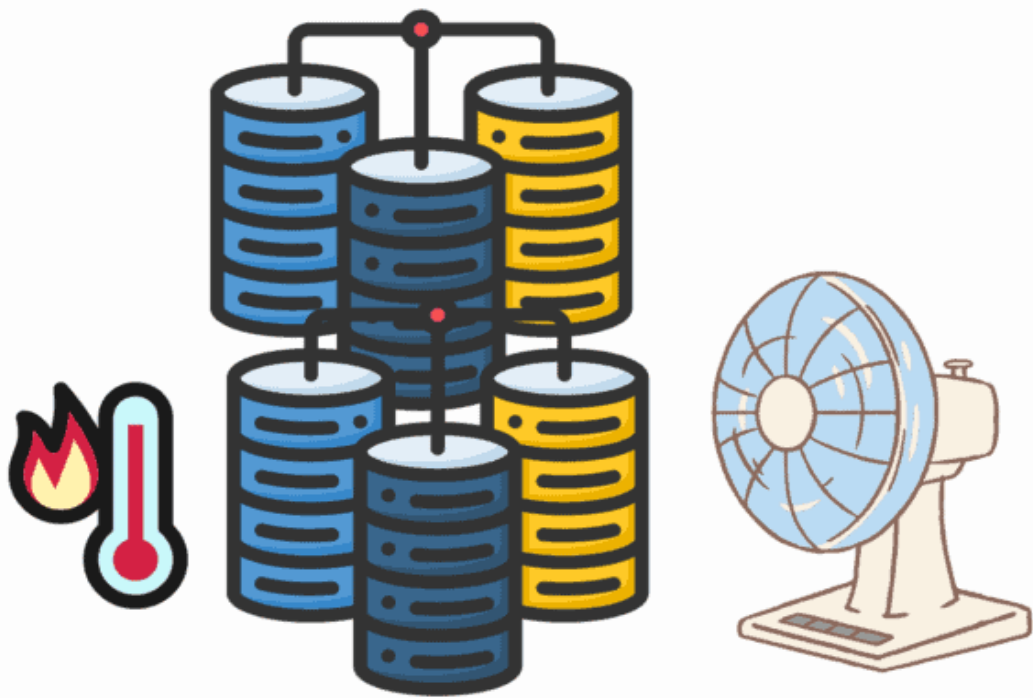
Large-scale data centers have drawn criticism for the amount of energy they consume. Running that many computers continuously generates a lot of heat, so air-conditioning is required. The cooling systems also require a lot of energy to run.

Location, such as placing data centers in moderate or colder climates, can reduce some of the energy bills. Unfortunately, proximity often improves response times, and it is not always viable to place data centers in the most power-optimal locations. Infrastructure to support such projects is also a consideration.

Many modern AI applications create additional power draws on the data center infrastructure. This is because their power consumption is significantly higher than traditional algorithms:

If that power use is going towards medical research or high-impact applications, that may be acceptable. Currently, however, many of the applications being served are not high-impact.

For example, [generative AI models](#) are being applied to memes and other art-based projects. That said, many of those [applications](#) are building on and helping to develop further technological improvements, which may provide unknown benefits.



4.2. Data Ownership

This topic has drawn a lot of discussion recently due to the proliferation of large image-generator methods. The current success stories are diffusion models. As with all large neural network-based models, a lot of data was required to train the models. This data had to come from somewhere.

It was collected from across the internet. It was often used without the express consent of the creator. The generated images can have a striking resemblance to the original. This setup has left a lot of artists and content creators angry as they believe their work has been stolen and used to create a tool to replace them.

This discussion of ownership and copyright is an important one. One need look no further than the recent Hollywood writers and actors strikes. Many companies are currently indiscriminately collecting the data they need to train their models. These legal questions highlight the importance of the ethical issues.

Building trust with an entity you believe is stealing from you is difficult. Maintaining data sources, provenance, and chain of custody is important in answering these challenges.

4.3. Existential Risk

Much has been said about existential risk, with arguments being made by serious AI researchers on both sides of the fence. A good place to start would be Yoshua Bengio's blog post on AI risk with questions and answers.

4.4. Consciousness

The problem of Consciousness is a long-standing one; and beyond the scope of this introduction. Those interested may find the article *What is it like to be a bat?* of interest.

A less technical read dealing with the same concept is the sci-fi story "For a breath, I tarry".

5. Conclusion

In this article, we discussed many practical and theoretical considerations when dealing with AI ethics. Many practical considerations are relevant now as AI systems become more ubiquitous across the tools we use. It is important that people can trust the technology and that the systems put in place don't unintentionally entrench existing biases or cause new harm.

Addressing these ethical considerations often involves balancing the principles of justice, fairness, human rights, and sustainability. It requires international cooperation, policy frameworks that consider the needs of vulnerable populations, and a commitment to sustainable development and the preservation of ecosystems for future generations.