Responsible AI

Modern AI: Challenges and Principles

Why worry about Responsible AI:
1.Increasing inequality
2.Weaponization
3.Unintentional Bias
4.Adversarial Attacks
5.Killer Drones
6.Deep Fakes
7.Data Poisoning
8.Hype

Microsoft AI Principles

Six Principles Guiding Microsoft Responsible AI Development:

1.**Fairness**:AI systems should treat all people fairly

2.**Reliability and Safety**:AI systems should perform consistently and minimize risk.

3.**Privacy and Security**:AI systems should protect people and their personal data.

4.**Inclusiveness**:AI systems should empower everyone and engage people

5.**Transperancy**:AI systems should be understandable

6.**Accountability**:Algorithms and the people who write them should be responsible or answerable for their impacts.

Model Transparency and Explainability

Model explainability is one of the most important problems in machine learning today. It's often the case that certain "black box" models such as deep neural networks are deployed to production and are running critical systems from everything in your workplace security cameras to your smartphone

Model Explainability can be grouped into two parts:

1.Direct Explainers:

- Model-specific Direct Explainers:SHAP Tree Explainers
  SHAP Deep Explainers
- Model-Agnostic Direct Explainers:Mimic Explainers
  SHAP Kernel Explainers

2.Meta Explainers:

- Tabular Explainer
- Text Explainer
- Mimic Explainer

Model Fairness

In machine learning, a given algorithm is said to be **fair**, or to have **fairness**, if its results are independent of given variables, especially those considered

sensitive, such as the traits of individuals which should not correlate with the outcome.

Fairlearn: Toolkit to identify and mitigate unfairness in machine learning models

Basic Principle of Group Fairness: Which group of individuals are at risk for experiencing harms?

Below, in yellow, are some ways ML fairness can be applied at various stages of your model development: