MODULE : 4

Responsible AI

The principle of developing and using AI ethically, with the intent of benefiting people and society while avoiding harm

Systemic bias

A tendency upheld by institutions that favors or disadvantages certain outcomes or groups

Data bias

A circumstance in which systemic errors or prejudices lead to unfair or inaccurate information, resulting in biased outputs

Allocative harm

Wrongdoing that occurs when an AI system's use or behavior withholds opportunities, resources, or information in domains that affect a person's well-being

Being aware of potential harm and negative outcomes is a first step to using AI responsibly

Data biases are circumstances in which systemic errors or prejudices lead to unfair or inaccurate information, resulting in biased outputs. Using AI responsibly and being aware of AI's potential biases can help you avoid these kinds of harms.

Phenomenon that can cause unfair or inaccurate outputs is drift. **Drift** is the decline in an AI model's accuracy in predictions due to changes over time that aren't reflected in the training data. This is commonly caused by **knowledge cutoff,** the concept that a model is trained at a specific point in time, so it doesn't have any knowledge of events or information after that date.