IBM AI Ethics(mod1)

Module 1

What is AI ethics?

The age of AI

Artificial intelligence is everywhere. It powers the navigation apps that help you find the most efficient or most eco-friendly route. It drives the search engines that help you find the most relevant information. It helps doctors reach more accurate diagnoses and develop more optimal treatment plans. It improves weather forecasting, enabling you to better prepare for significant weather events. In conjunction with sensors and satellites, it can collect data about the environment that helps scientists better understand and make predictions about our changing world. AI can make life easier and safer by helping people make more informed decisions, connecting them with the right information they need at the right time, and finding patterns or efficiencies that they might not otherwise know about.

But AI also has the potential to do harm. For example, AI can be used when determining who gets a loan, who gets accepted to college or selected for a job, how employees are compensated, or even the lengths of prison sentences. In the context of AI, harm doesn't necessarily have to do with physical harm. The harm can be less

obvious, taking the form of inequity, discrimination, or exclusion. And this harm can be subtle because people may not always know when they are interacting with AI or when and how AI may be influencing decisions about them.

Can AI be trusted?

As AI is increasingly embedded in everyday life, it is vital that people can trust AI. Practitioners infuse trust into AI systems with **AI ethics**. AI ethics is a multidisciplinary field that investigates how to optimize AI's beneficial impact while reducing unintended or adverse outcomes. There are five pillars of AI ethics: fairness, robustness, explainability, transparency, and privacy. These pillars are focus areas that help make AI trustworthy.

In the following video excerpt, John Thomas, a Vice President and Distinguished Engineer at IBM, explains more about each of the five pillars.

> this part is uploaded on github

AI is a part of our lives, and its influence is growing, so it is crucial that people who design, develop, deploy, procure, and use AI understand how to use AI ethics to minimize harm and optimize benefits.