

THE ETHICS OF AI

VIDEO TRANSCRIPT

Imagine you oversee the issuing of government loans designed to help citizens with housing needs. You develop an artificial intelligence algorithm to predict whether prospective borrowers will pay back their debts. You don't know how it does it, but the system works – weeding out bad debts, and loan defaults falls significantly.

But is it okay that you don't understand how the system is making the decisions? What if there is something unfair or prejudiced in how it works?

This is an example of the ethical puzzles emerging from the use of AI. The challenge as with any new technology is to make sure it is used responsibly.

We also must decide which ethical choices we can leave to AI systems as different people think differently about what is ethical or unethical. Unfortunately there have already been real examples of ethical failures in AI, from sexist resumes screening by recruiters to racist risk assessments in judicial systems. Blackbox AI systems are not designed to explain how they make decisions. They are trained on vast amounts of data and find patterns to deliver the desired outcome.

What is needed is another application to work as an observer analyzing the main system to decide if it is ethically sound. It is a challenge to decide what is right and wrong in a fast-evolving field, but using the guiding principles of accountability, transparency, fairness, and honesty are a must.