

Discussion Guide: Ch 10, AI and Ethics¹

Goal: Explore how society should manage AI and who is responsible when things go wrong.

Topic 1: The "Timing" Problem

Policy usually happens **after** technology is already being used by millions of people.

- **Discussion Question:** If we regulate AI before it is fully developed, do we risk "killing" innovation? Or is it more dangerous to wait until the AI is already "out in the wild"?

Topic 2: Who Gets the Blame?

When an AI system causes harm (like a biased hiring algorithm or a self-driving car accident), who should be held accountable?

- **The Options:** The government for not regulating it, the company that sold it, the engineer who wrote the code, or the AI itself?
- **The "Electronic Person":** Some propose giving AI a "legal personality" (like a corporation) to handle liability. Others think this is a "dangerous idea" that lets humans off the hook.

Topic 3: The "Ethical Black Box"

Some researchers suggest every autonomous AI should have an "ethical black box" that records its decisions, similar to an airplane's flight recorder.

- **Discussion Question:** Does knowing *why* an AI made a mistake make you trust it more, even if the mistake still happened?

¹ AI helped generate this page, with instructor checks.