PS643

3rd September 2024

The Chilling Effect: Adopting policies that create fear among people, pushing them to be more disciplined automatically.

- Cameras on the road—some don't even work, but people fear them.

- It's less about productivity, more about employees feeling terrorized into compliance.

Snake Oil: Using fear/greed to sell anything.

Prediction of social outcomes: criminal recidivism, job performance, terrorism

COMPAS tool: Initially designed to study prisoners' characteristics and identify those needing mental health support, uses a complex set of features. But a simple logistic regression model with just two features(age,the number of cases)—gave better accuracy in predicting risk.

AI Divisions: Perception, Automating Judgment, Predicting social outcomes.

- Very simplistic division, but helps understand the scope.

- **Recidivism**: Whether a criminal will reoffend.

- You can hold a judge accountable for bias, but not AI—even if AI has 99% accuracy, that 1% margin of error can't wrongly convict innocent people.

Harms of AI in Predicting Social Outcomes:

1. Endless hunger for personal data

2. Lack of explainability

3. Massive power shift from experts to tech companies, who remain unaccountable

4. Creates the illusion that a problem is solved, when it might not be

Pillars of Philosophy

- **Ontology**

- **Epistemology**

- **Ethics**

What is ethics, are ethics and morality the same?

- **Metaethics** (What is ethics?)
- **Normative ethics** (What's the right thing to do?)
- **Applied ethics** (How to apply ethics in real life?)

Virtue ethics (oldest): Not about actions performed by the person, just character.

Deontological Ethics: intention behind the action?

- good/bad.

Immanuel Kant — Categorical Imperative.

Consequentialism: impact that the action had on the society.

impact helps decide if it's right/wrong

Utilitarianism: Who decides what's useful? Is the state biased?

Applied Ethics:

- **Policymaking**: Every policy needs a clear framework.

- **Bioethics and machine ethics**: Should machines help in ethical decisions?

- **AI control problems**: What issues arise from AI's involvement?

- **Effective altruism** : Gather as much wealth using AI, donate it