

## AI Snake Oil. (Paper #7)

- AI can't predict / quantify internal capability. (eg. thought process).
- Only observable things can be predicted.
- Companies still making money even if fraud because:
  - Inspires people to react a certain way.
  - Create fear so that people pretend that prob. doesn't exist, & forcibly modify behaviour
  - Makes employer look diligent.
- Nobody (except AI scientists) cares whether it works.
- Employees being exploited → actual reason for seemingly improved performance.
- This is not productivity, which is same i/p, ↑ o/p.  
But what happens here is ↑ i/p ⇒ ↑ o/p.
- 2029 prediction: Incentivizes people to put money in harmful things rather than actually useful things.
- 3 types of AI systems
 

↓

Perception

↓

works

↓

Automated judgement

↓

Conceptually acceptable but doesn't always work

⇒ *dicey*

→ Social outcomes

↓

Can never work.

↓

*human Psychology cannot be predicted.*
- Society complex / fundamentally pseudo-scientific → reason of dubiousness in 3<sup>rd</sup> category.  
*most dangerous type.*
- COMPAS to predict recidivism.
  - doesn't work + biased (racism).
- Even if 99% accuracy, not acceptable since innocent shouldn't be punished.
- Even if human error → moral responsibility.
- Most social outcomes operating with legislature than judiciary.

## Slides week 6

- Philosophy → at the base of everything.
- Basic pillars of philosophy
  - ① Ontology: what is reality?
  - ② Epistemology: what is truth? what is knowledge?
  - ③ Ethics: moral philosophy.

## Mind & Ethics

Q1. What is ethical thing to do → *Normative ethics*

Q2. What is ethics → *Meta ethics*

Q3. How do I translate it to real life (eg, AI) → *Applied ethics.*

*→ obsolete now.*

- Old times: virtue ethics → does a **person** have good qualities.  
→ Hero / villain only.
- Deontological ethics: intention behind an **action**.  
⇒ Immanuel Kant → categorical imperative (regardless of reason, if action is wrong, it is not acceptable).
- Remove biases, child-like mind, ignorance, impartial } *rights*  
→ to tell which action is good or bad
- Consequentialism: how action impacts society  
→ utilitarianism.  
⇒ Bentham → universalism, contextualization.
- Pragmatism: American  
→ representation  
⇒ mutually agreeable.
- Law → deontological.  
Constitution → pragmatic.