

03/09

Page No.:

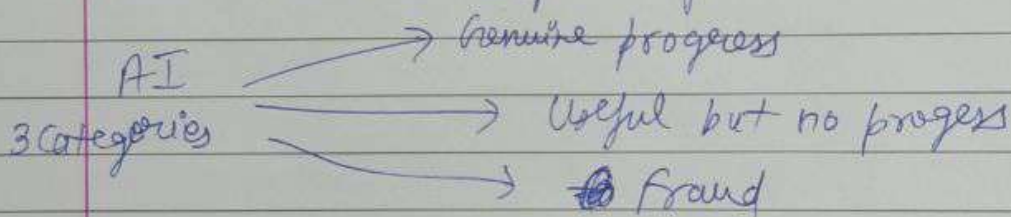
YOUVA

Date:

- ① UI of an AI app^{or company} that can send job recommendations to people. Why apps like this fraud?

Because No AI can tell you what you actually are from inside, it's not something that can be predicted, as it is not observable. Things like how a person thinks/is suitable for is not ~~is~~ predictable by AI.

Company's like this are funded by millions of \$s. Because using this company's can get idea of there employees and hires, although they think of such tools like a ~~magic~~ magic box.



- ① Genuine progress → eg: face recognition, it works, speech to text works, deepfakes works
- ② AI is liey → judgement ~~task~~ tasks
- ③ fundamentally ~~does not~~ work → social outcomes decisions.

→ Princeton study: to get a list of scientist and get their opinion about welfare of child. They found ~12000 variables to measure. Then they found some factors which can predict whether a child will turn out to be good/successful. Result → 13000 features were hardly better than only 4 features.

America very prison heavy country.

Bail → you are under trial

prison → you have been found guilty

People at prison having mental health issue, then they decided to help only some very affected ones, by using AI to find such prisoners.

Can it also predict which prisoners are most dangerous?

They decided to use COMPAS tool for bail of prisoners or parole.

Recidivism : means whether a criminal will reoffend, that is do the same crime twice.

But, just using age & no. of prisoner factors with logistic regression ~~was~~ turned out almost same result as COMPAS. COMPAS was biased!
↳ Black/white

Claim of prof. Narayan : For predicting social outcomes, AI is not so good.

→ Take aways: AI not good on social outcomes. In most cases, manual scoring almost similar accurate.

Week 6 slides

F/A/T, Ethics, Human Rights

Philosophy : meta area of scholarship which deals with what reality is, what human are etc.
Eg : Science was called Natural Philosophy.

Economics is part of moral philosophy.
 # ~~That's not~~ Every thing is part of philosophy,
 that's why the biggest degree you get is
 Phd !!

Thinking → ~~It's not~~ ? is philosophy

→ Questioning of

- Q1 → What is all this? { Ontology } reality
 Q2 → How do I know what is right (truth)? { Epistemology }
 Q3 → What do I need to do? { Ethics, moral philo }

Ethics → (1) What is ethical things to do? (Normative)
 (2) What is ethics? (Meta)
 (3) How do I ~~use~~ that in real life? (Applied ethics)

- (i) ~~Old~~ Old thing: Virtue Ethics, boils down to asking that does a ~~person~~ person have good qualities?
 (ii) Deontological: ~~Intention~~ Intention behind actions?
 ↳ "Immanuel Kant" German philosopher
 John Rolles: Will of ignorance, you must remove all your bias/desires etc to judge something, that it is good or bad.
 But deontological did not look at cause or consequences.
 (iii) Consequentialism: Only look at what action impacts society as. Good is what makes people happy.
 What is utility of your action.
 (iv) Pragmatism: Get everybody on table & find something which is ~~mutual~~ mutually agreed on.
 This is by Americans.

PS643 (06/09/2024)

Atishay Jain

Meta Ethics

- Applied ethics: ability to convert your conclusions into real life
- Ethical statements
- Moral statements - some think they are an attempt to describe reality and so have a true/false answer (such people are called cognitivists people), others who think the opposite are called non-cognitivists
- If moral statements are true/false answer types, then is any moral statement true?
- Cognitivists
 - Some think that every moral statement is FALSE and is an error - called Error theorist people
 - Those who think that 'some' moral statements can be true are called Moral Realists (in an objective sense)
 - Those who think that 'some' moral statements can be true in subjective/social/cultural context are called Moral subjectivist
- Non-cognitivists
 - Moral statements are Emotional statements - emotivists
 - They are orders, people think about them on their will - prescriptivism
 - While not real, humanity treats them in a quasi-real way to live a life - Quasi realism
 - Is-ought gap: easy to describe reality when we talk about logic but it becomes difficult when we start saying "ought" or "should have". Like Humans ought to live nicely, Humans should have justice. Where does this ought come from? Where does these intentions come from? Because what you want from reality you cannot describe in descriptive statements, at some point you think that it should be like this.
- Ethics - biggest problem that it is impossible to impose

Project Maven

- We want to have entire biodata of any person on earth at anytime, using satellites to look at them
- Everything about you just by seeing you via a satellite, pentagon wanted this
- They gave this task to Google, and they did something, and finally engineers decided to strike and get fired
- When google realized that it is an unsolvable problem, then they told SORRY to US government
- Moral of the story:
 - When it comes to ethics, we start talking about law & regulations, but we forget that humans make them and humans have some moral actions
 - Those employees decided not to make such a thing rather get fired 🔥
 - Its the employees which make the AI

Lack of Transparency

- Transparency of ethics councils

- How ethics decided
- Everything from bulb making to AI, bureau of indian standards makes standards for them
- People who make these are neither democratically chosen nor by some means/exams
- They are professional but how these bodies are formed is not very transparent

Issues with tech Ethics

- Individual solutions. Problem remains. The problem is structural here, it is that how can the American govt even ask for such a project to companies, its unethical.
- Corporate bodies moderate them, which are supplied by money
- Ethics are vague and contradictory
- People think tech can solve social problems
- Determinism is the belief that history is determined by tech
- Ethicswashing - deliberate and malicious use of ethical logic in order to divert the discussion or not do anything real

What is a mind?

- When you sleep, you not conscious
- Nobody knows how consciousness works
- Chinese Room
 - Imagine there is a room here with a small window
 - From that window, you can only just send a A4 sheet
 - There is a person inside the room, and he don't know chinese
 - But he has a magical dictionary inside the room which has all chinese words
 - From outside, I write something in chinese and send inside the room
 - The person writes its meaning in english and gives me back
 - Technically chinese -> english translation happened, and I may think that the person inside the room knows chinese
 - I know both, but inside the room only symbols have been changed
 - We think it is intelligent, but the neural network can't do any reading, it just plays with symbols/math
 - Neural network doesn't know what its doing, humans make use of its input/output
 - Hard AI
- Dualism
 - Mind and the body/matter are separate
- Monism
 - Mind is not separate from matter, it has a scientific explanation
 - Idealists: only reasoning can be happened in form of ideas
 - Materialists : believing that reality outside mind can be measured