Sept-3

→ Snake-oil reading:-

AI → personality assessment } → fraud:
↳ no AI can tell you what a person is or suitable/not suitable
prediction } only what is observable.
thinking & suitability } are not observable } not tangibles.

→ funding? → I would pay good money if it inspires terror in my employees.

* "chilling effect" → (enacting certain policies not with the agenda of solving anything but instead scares into discipline)

"SK → companies recommend plastic surgeries & how to smile so that office AI does not target them"

→ (AI on twitter)
Software increases productivity X → terror makes us work hard
→ productivity in econ :- same input → more output
here more " ⟷ more "

→ Why so much snake oil in AI?-
↳ In older America — people claim magically medicine.
Snake oil → selling using fear & greed of people

* public opinion:-
↳ thinks AI can be powerfull } dangerous
↓
incentivises AI products even though not usefull.

AI works & well } genuine progress.
AI works - part bad } carefull. → face recognition
AI fraud & harmfull (deep fakes
Speech to text)
works

no interpretation / reasoning { harmfull
only identification
automatic essay grading.

don't work :- Spam detection, Content recommendation
but improving ↳ not always good performing

ethical concerns → high accuracy.
→ errors

. fundamentally dubious :- deciding social outcome / involving telepath... or predicting human beha...
emotion detections } doesn't work.

↓ don't work
→ complex society
→ fundamentally pseudo scientific priors.

→ fragile families :- list of features to predict child well being
457 scientists } 12,942 variables. for 4,242 fam...

results ≅ with 4 features
↓ linear regression

* COMPAS :- } problem → too many people in jail
jail vs prison → guilty
↓ undertrail

↳ lot need mental help
↳ ~~help all~~ ↳ help most victimised people wi... AI ? } compas tool.

predicts which prisnors is most
vulnerable
~~guilty~~. dangerous → (de... after decid...

process is punishment }?

COMPAS for bail/parol or recidivism (prob of reoffending)
↓ doesn't work but America uses it
↓ giving parol/not

→ fairness :- COMPAS → 65% ± 1%.
2-featured logistic regression → 67% ± 2%.

also biased.

→ would you use it if it has 99%.? → No
better 100 guilty than 1 innocent punished.

→ COMPAS has no morality.

Harms:-

Smart city in India } certain things use digital automated solns given by some companies.

→ problem not actually solved but seems to everyone.

people can't vote these & they're unaccountable.

paper 5 reading:- bird's eye view of how people think about AI ethics.

Week 6 slides:- F/A/T → Transparency
fairness, accountability

→ philosophy to understand ethics.
(PS 144) deals with what reality is? what humans are .. etc

Science = Natural philosophy

Moral philosophy } how human's should act
(economics is a part)

philosophy encompasses everything } literal statement not rhetorical

→ pillars of philosophy

ontology (what is reality?)

how do I know right/truth (Epistemology)

what do we need to do? (Ethics/Moral philosophy)

(Mind & Ethics)

→ In common world: morality → religious, ethics → secular } × to philosophers it's same.

→ questions.

1. How do we decide it is ethical } Normative ethics
2. What is ethics } Meta ethics.
3. How do I translate to real life ethics } Applied ethics.

Normative :- what's right thing to do?

→ how people decide right/wrong } not consistent, widely changed over time & space

(a) ancient greek/china/India } (Virtue ethics) individual

(does a (person) have good qualities?) doesn't look at actions.

Ex:- Hercules in greek → burns countries but is hero.

(b) (Deontologicial) ethics } → what was the intention behind the action.

↓ looks at fund intention of action → good/bad.

"The critique of pure reason" } → categorical imperative
↓ limit
*(Emmanuel...)

↓ doesn't care about the situation. Ex:- Murder is bad

* John rogs } → Veil of ignorance

first remove all context & bias (desire... etc) & be clean } then we can tell if action is good/bad

Natural rights } criticism → didn't look at cause & consequence of action.

③ Consequentialism: → only look at the impact of action.

Bntham } → what is good? → what makes people happy

utility of action

problem:- who decide utility

↓

State consequentialism

problem:- ↓ if state is partial.

Carl Max } → what you call universal utility is } shopkuper mentality.

for 1 shopkuper (bntun)

↳ still a consequalist

Welfarism } improve life of people.

④ Pragmatic:- → very american way of ethics.

→ get everybody of different social background to decide mutual.

Indian law } deontological

a constitution } pragmatic

Applied Ethics:-

↳ Machine ethics }? can machines be moral agents?

AI control problem:- how do we make AI's which don't do what we don't want it do?

→ is this a real problem? why does this have control over others?

effective

Altruism & longtermism:-

↑ use AI to make as much money & distribute to people best ?