

Metaethics

Policy: whether argument is right or wrong - normative ethics

- brief given to policymaker - applied ethics
- what is ethics?

Description of nature: earth is flat, sun goes around earth
(Does not matter if accurate)

Moral statement → attempt to describe reality

↓
Cognitivists (true/false answer to statement)

Those who believe otherwise and thus moral statements do not have true/false answers are non-cognitivists.

- Error theorists: believe that every moral statement is false
- Moral realists: some moral statements are true objectively
- Moral subjectivists: subjectively true moral statements, only in a socio-cultural context, not in absolute context

Moral non-cognitivists

Emotists: morality is emotional

Prescriptivists: orders or wills of nature

Quasi-realists: morality is quasi-real, made by humans to live life

Is-ought gap: David Hume

↓
Gap between description of reality and desire (reality should be)

It is impossible to enforce → not everyone agrees on ethics

→ not actionable as there is no consequence

- Bureau of Indian Standards - self-replicating, no exam/election
same with other bodies like IEEE
formation of body is not transparent

Project Maven - Google employees strike - govt. project

Solutionism - people think tech can solve social problems

Determinism - history decided by tech

Chinese room - person inside does not know Chinese but has magic Chinese dictionary. Slide a paper in, in Chinese, person uses dictionary and sends back English paper

Person ~~inside~~^{outside} thinks translation occurred, but no, just used symbols

- Analogy to neural networks

↳ only does symbol flipping, not intelligent

Dualism: body and mind separate

Monism: mind and matter are connected

↳ Idealists: cannot think about matter as it requires mind → only ideas

Materialists: reality outside mind can be measured

Free will: if monism is true, matter is controlled by laws of physics and connected to mind, is mind controlled by laws of physics? does that mean all thinking is pre-programmed?

→ Behaviourism: mind expresses itself as behaviour

↳ wrong, medical negative consequences like lobotomy

Blackbox: thoughts and brain connection

Qualia: people may perceive colours differently (eg.)

Debiasing

- Sampling: Data represents the population accurately
- Model and output level: No bias against protected groups (gender, religion, etc.)
- Explainability - irrelevant to end user, relevant to engineer
procedural vs technical explanation

Issues with FAT

- McNamara fallacy

↓
anything that can't be measured is not knowledge and thus must be ignored

Qualitative vs quantitative knowledge

↓
surveys

↓
CS

- Epistemic contestation

Epistemology - how to know what is right

eg: lawyer says no to facial recognition, engineer says yes due to high accuracy - different viewpoints.