Ethics, Philosphy and AI

Martín Pereira Fariña

Universidade de Santiago de Compostela

Department of Philosophy and Anthropology

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA   FACULTADE DE FILOSOFÍA

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA   DEPARTAMENTO DE FILOSOFÍA E ANTROPOLOXÍA

Outline

- How philosophy of language and AI are related?

- How ethics and AI are related?

- How political philosophy and AI are related?

What is meaning?

- Semantic theories of meaning

- Pragmatic theories of meaning

What is meaning?

- Semantic theories of meaning
    - What is the meaning of this expresión?
    - Content and reference
    - Description of the world: truth-value
    - Frege, Russell, Wittgenstain (1st)
    - Frege: "*The **reference** of a proper name is the **object itself** which we designate by its means; the idea, which we have in that case, is wholly subjective; in between lies the **sense**, which is indeed no longer **subjective like the idea, but is yet not the object itself**.*" ((Frege 1892 [1960])

What is meaning?

- Pragmatic theories of meaning
    - Under what circumstances an expression has a meaning?
    - Communicative intention
    - Achieving a communicative goal: felicity conditions
    - Wittgenstein (2nd), Austin, Searle, Grice
    - Wittgenstein: "*For a large class of cases of the employment of the word 'meaning'—though not for all—this word can be explained in this way: **the meaning of a word is its use in the language**"* (PI 43, Wittgenstein, 1953)

- The argumentative theory of mind: i) justifying ourselves; ii) persuading others. (Mercier & Sperber, 2018)

How can we understand an AI-generated linguistic explanation?

- Lucie's application mortgage rejected (Cappelen, 2020)

    1. What does the output '550' that has been assigned to me mean?
    2. Why is the '550' that the computer displays on the screen an assessment of my credit-worthiness? What makes it mean that?
    3. How is the final meaningful state of SmartCredit (the output '550', meaning that Lucie's credit score is 550) the result of other meaningful considerations that SmartCredit is taking into account?

How we can understand an AI generated linguistic explanation?

- Semantic approach: **no meaning**
  - Chinese room argument (Searle, 1980).
  - The AI system needs content, reference and produce statements to have truth-values.
  - Content doesn't rely on the internal computational structure in the same way that it doesn't rely on the internal strcture of our brain.
- Pragmatic approach: **it might have a meaning**
  - Communicative intention: reasons for actions (normative reasons, motivating reasons and explanaroty reasons)
  - Goal: An AI system should generate explanatory or motivating reasons for human -> externalism, argumentative.
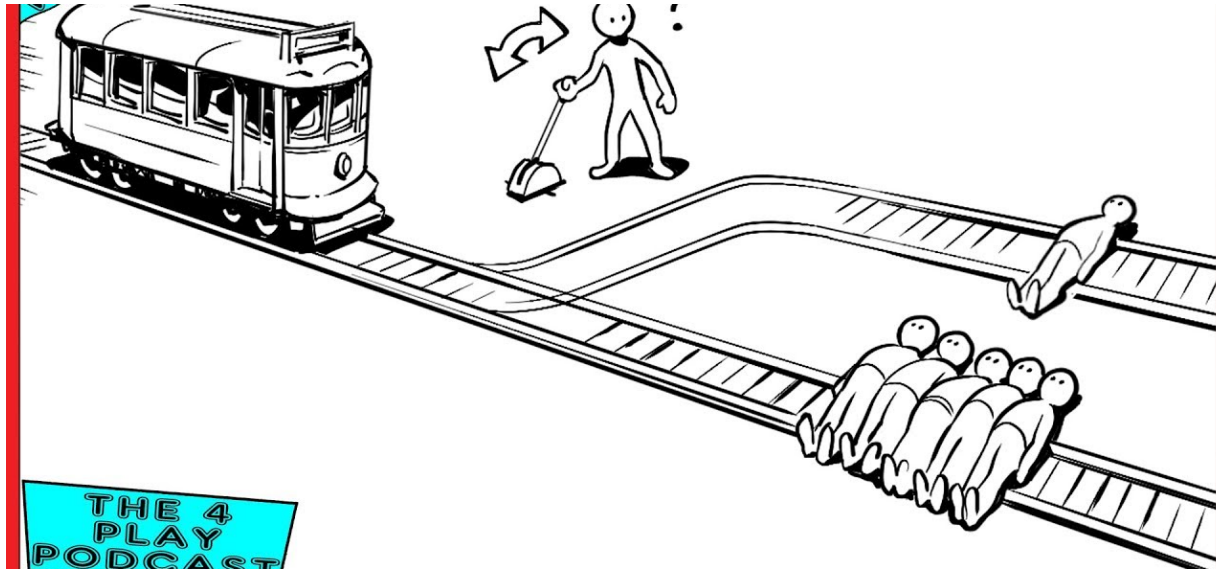
How ethics and AI are related?

- Two types of relations (Coeckelbergh, 2021):

    - *AI ethics*: how humans should use, develop and interact with AI systems that leads to good things for humanity.

    - *Ethics for AI*: the "moral module" that can be inserted into an AI system.

(Meta)Ethics and AI

- "Respect ethical principles": which ones?

- Consequentialism: Moral actions are evaluted according its consequences (John Stuart Mill, utilitarism).

- Deontology: Moral actions are evaluted according to moral universal laws (Kant).

- Virtue ethics: Moral actions are evaluated by their contribution to the development of good character virtues (Aristotle); human flourishing (Carston, 2020).

Moral dilemmas

Towards a Virtuous AI

1. Human flourishing is central to the development of AI.

2. Humans as social animals can only fourish in society.

3. Flourishing requires humans to do what we are specially equipped to do.

4. We need to acquiere genine knowledge via theoretical reasoning and then act autonomously and justly via practical reasoning in order to flourish.

5. The key to excellent practical reasoning and hence to being ethical is the ability to deliberate about one's goals and choose a wise course of action.

(Bynum, 2006)

How political philosophy and AI are related?

- Analysis and comprehension of production relations.

- *Work is one of the main mechanisms for the redistribution of wealth ->* this not as clear now.

- How AI is affecting these type of relations?

Amazon go

Culture

ISABEL RUBIO

Seattle - NOV 01, 2023 - 20:45 CET

AMAZON >

This is what Amazon Go's self-service supermarkets with no cashiers are like

EL PAÍS visits a Seattle store where multiple cameras detect what produ customers are taking with them

store located in downtown Seattle, on the corner of Terry Avenue and Stewart Street. The supermarket is only open on weekdays, from 7 a.m. to 8 p.m. With only three aisles, it feels small compared to other American supermarkets. The ceiling is covered with cameras that "identify individual customers and record the items they select," said a store employee. Amazon doesn't provide much information about the system, but says it uses computer vision, sensors and deep learning techniques. "When you grab something off the shelf, it gets added to your virtual cart automatically. And if you decide to put it back, it gets removed from the virtual cart too."

Amazon go

Amazon Fresh kills "Just Walk Out" shopping tech—it never really worked

"AI" checkout was actually powered by 1,000 human video reviewers in India.

RON AMADEO - 4/3/2024, 6:55 PM

Ghost work

- Workers that solve the things that AI can't do: new global underclass (Mary & Suri, 2019).
- Vicious
  - No social protection
  - Anonymous
  - Outsorcer labour
  - No deparments to make complaints
- Virtuous
  - Incomes for remote áreas
  - Salaries for women that can't work outside home (Asia)
  - Self organisation

Conclusions/Claims

- Explainable AI is also a matter of provide justifications and arguments for actions.

- Ethics of AI is not just a matter of embedding ethics into machines and aim for a more virtuous society, not just avoid harm.

- Economy of AI is not just about creating or destroying Jobs, it's also about the how the labour relations are modified.