

# What can philosophy tell us?

Tim Crane, CEU Vienna, Department of Philosophy

A Paradigm Shift in Computer Science? TU Wien 28-29 November 2024

# Two philosophical themes

1. The question of intelligence
2. The question of ethics

# 1. The question of intelligence

# AI and AGI

To create genuine 'thinking machines', we must create 'Artificial General Intelligence' (AGI)

Not just special purpose machines like Deep Blue, AlphaGO, ChatGPT etc

But 'general intelligence' as we have it



Hubert L. Dreyfus

# What Computers Can't Do

REVISED  
EDITION

THE LIMITS OF  
ARTIFICIAL INTELLIGENCE

# Philosophers on AGI

“Artificial general intelligence is possible ... There are a lot of mountains we need to climb before we get to human-level AGI. That said, I think it’s going to be possible eventually, say in the 40-to-100-year time frame”

David Chalmers, *New York Times* 2019

# Scientific and conceptual questions

- Questions about possibility and essence are philosophical questions
- In this case: the question about the essence or nature of intelligence

# Generalising from recent achievements

- ‘We’ve done GO/ChatGPT/DALL-E (etc) so now all we need to do is AGI’
- But GO is still a **game**
- And a game must have a goal



# Goals

- Face recognition, text recognition, voice recognition, responding to a prompt....
- They all have goals

# The upshot

‘If intelligence is a problem-solving algorithm, then it can only be understood with respect to a specific problem’

François Chollet, ‘The Implausibility of Intelligence Explosion’ (2017)

# The goal is essential

- If there is no goal, then there would be no such thing as succeeding or failing in achieving it
- So there would be no such thing as training a machine to achieve it, or telling it when it has failed

# What is the goal of general intelligence?

- No-one in AI has any idea
- Is this because it is a deep and difficult question, needing more empirical investigation?
- Or is it because it is a bad question?

## 2. The question of ethics

# What are the ethical questions about AI?

‘All politicians must understand the difference between artificial intelligence and artificial consciousness. The unintended or even intentional creation of artificial consciousness is highly problematic from an ethical perspective, because it may lead to artificial suffering and a consciously experienced sense of self in autonomous, intelligent systems’

Thomas Metzinger, ‘Should we fear AI?’ European Parliamentary Research Service, 2018

# The question of how we treat AI systems

‘Consciousness matters morally. Conscious systems have moral status. If fish are conscious, it matters how we treat them. They’re within the moral circle. If at some point AI systems become conscious, they’ll also be within the moral circle, and it will matter how we treat them. More generally, conscious AI will be a step on the path to human level artificial general intelligence. It will be a major step that we shouldn’t take unreflectively or unknowingly’

David Chalmers, ‘Are large language models conscious?’ *Boston Review* 2023

# Why these questions?

- The question of the extent of the ‘moral circle’ is important
- But why think that it will ever apply to computers?



# Creating artificial consciousness

‘If you were to build a computer that has the same circuitry as the brain, this computer would also have consciousness associated with it’

Christof Koch, *MIT Technology Review* 2014

# Assumptions, further questions

- Everything that the brain does/produces is computational
- Since the brain produces consciousness, the brain processes producing consciousness must have a goal
- Is consciousness a mere side-effect of these processes?
- Or is consciousness itself the goal? What does this mean?

# Other more concrete problems

- Excessive reliance on mechanisms which are epistemically unreliable
- Privacy
- Cybersecurity
- Deepfakes, plagiarism
- Environmental costs

# Conclusions

- (1) AI raises many urgent philosophical questions about ethics and intelligence
- (2) But a lot of philosophical thinking concentrates on the wrong questions
- (3) These questions involve large and controversial assumptions: worth discussing but not urgent
- (4) We should concentrate on the philosophical questions posed by 'real existing AI'

# Thank you

[www.timcrane.com](http://www.timcrane.com)

**KNOWLEDGE**  
**in** An FWF Cluster  
of Excellence **Crisis**

