Course
Philosophy of Intelligence (人工智能哲学), Spring Semester 2021

Teacher
Sebastian Sunday Grève (ssg@pku.edu.cn) (office hours: Mondays 14:00–16:00)

Location
Weeks 1–16, Thursdays, 15:10–18:00, room to be announced

Level
Open to all undergraduate and graduate (both master's and PhD) students; no prerequisites

Description
Artificial intelligence will radically transform our lives. By many accounts, it has already begun to do so. The rapid technological development in this area raises important questions that require careful philosophical attention. Can machines be made to learn? Can machines be made to think for themselves? Can machines be conscious? Can machines have subjective experience? Might machines deserve or require moral consideration? What measures should we take to direct the development of such machines? What will happen if machines' capacities come to exceed our own? Does the advent of super-intelligent machines pose an existential threat to humanity? This course will give a thorough introduction to these kinds of questions and possible answers to them. It will include an overview of the history of computer science and the basics of machine learning. We will also study seminal works by past and present philosophers.

Format
The course will be largely a mix of lectures and seminars. A small number of meetings will be organised as workshops, which will be entirely focused on acquiring or improving philosophical writing and dialogue skills. Students will receive prompt written feedback on each requirement that they fulfil by the end of Week 16.

Marking
Each student's individual marks for the various assessed requirements will be weighted such that the maximum percentage shown below is assigned to their highest score (see Requirements).

Requirements
Weekly reading assignments, usually one article or chapter
Participation (10–30% of the overall course mark)
2 × 500-word essays by Week 6 (10–20%; only the higher of the two scores will be counted)
1 × either 1,500-word essay by Week 12 or 20-minute presentation during term (20–40%)
1 × 2,500-word essay on an individual topic, at the end of term (30–50%)
1 × draft of the end-of-term essay, min. 1,000 words, by Week 14

Options
Students may revise and resubmit their 1,500-word essay once before the end of term, if they wish to improve their initial mark. In consultation with the teacher, students may also present work equivalent to one or more of the requirements set out above.

Selected readings

[This is a representative sample. All readings will be made available in electronic form.]

Boden, Margaret A. (ed.). 1990. *The Philosophy of Artificial Intelligence*. Oxford University Press.

Boden, Margaret A. 2016. *AI: Its Nature and Future*. Oxford University Press.

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Clark, Andy. 2001. *Natural-Born Cyborgs?* In Beynon, Meurig, Nehaniv, Chrystopher L., Dautenhahn, Kerstin (Eds), *Cognitive Technology: Instruments of Mind* (Springer), 17–24.

Copeland, Jack. 1993. *Artificial Intelligence: A Philosophical Introduction*. Blackwell.

Domingos, Pedro. 2015. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Penguin.

Dreyfus, Hubert L. 1992. *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT Press.

Kurzweil, Ray. 2006. *The Singularity is Near: When Humans Transcend Biology*. Penguin.

Russell, Stuart J., Norvig, Peter. 2010. *Artificial Intelligence: A Modern Approach*, third edition. Prentice Hall.

Searle, John R. 1980. 'Minds, Brains, and Programs', *The Behavioral and Brain Sciences* 3, 417–24.

Turing, A. M. 1950. 'Computing Machinery and Intelligence', *Mind* 59, 433–60.

Weekly schedule (to be confirmed)

Week 1: Introduction

Week 2: History of computer science and AI research

Week 3: Turing's mathematical logic, engineering, and philosophy of intelligence

Week 4: Searle's Chinese Room Argument

Week 5: Dreyfus's critique of AI

Week 6: Recent developments in machine learning research

Week 7: Andy Clark on embodied cognition and the future of human intelligence

Week 8: Nick Bostrom vs Ray Kurzweil on super-intelligence

The exact schedule of topics for weeks 9 to 16 will be determined later, depending on prior progress and how the students' interests develop. Areas covered will include machine consciousness, machine freedom, ethics of AI, and animal intelligence. Two or three meetings will be organised as workshops, with a focus on philosophical techniques and revisiting previous course content.