THE PHILOSOPHY OF ARTIFICIAL INTELLIGENCE

Time and Place: Tuesdays 16:15-17:45, 23.02/U1.22
Instructor: Dr. Ioannis Votsis
E-mail: votsis@phil-fak.uni-duesseldorf.de
Office hours (Room Geb. 23.21/04.86): Wednesdays 11:00-12:00

This course aims to probe the philosophical presuppositions that underlie discussions of artificial intelligence. We will start with a semi-historical introduction to several central issues and notions in the debate on artificial intelligence, e.g. notions of intelligence, the definition of an algorithm, and the Turing test. We will then proceed to examine to what extent the debate has moved forward since. For example, it was once argued that a computer would never be able to match humans in complicated reasoning tasks, including chess-playing at a world-class level. As is well known, computers have now surpassed that milestone. How much more can we expect computers to be able to achieve in the future? Will computers be able to think? If not, can their outputs at least be indistinguishable from those of human beings? What about consciousness? Will there ever be computers that are conscious and have human-like feelings? To take this line of reasoning to the limit, will there ever be computers that exceed humans in all these respects, i.e. that become ‘superhuman’? In pursuing answers to these and other related questions we will delve into discussions of behaviourism, consciousness, reductionism, intentionality, the Chinese room argument, simulation, neural networks, expert systems, the frame problem, etc.

Useful Anthology:

Coursework:
- One presentation (about 20 minutes) on one of the main readings. [3 credits]
- One essay (about 2,500 words), deadline 07/07/09. [3 credits]

NB: Presentations will be assigned on the second week. Suggested essay topics will be distributed in May.

WEEK 1: Introduction and Presentation Assignments

WEEK 2: Computability and the Turing Test

Main Reading:

Further Reading:
WEEK 3: A Modern Philosophical Theory of Mind

Main Reading:

Further Reading:

WEEK 4: Qualia

Main Reading:

Further Reading:

WEEK 5: Connectionism

Main Reading:

Further Reading:

WEEK 6: Classical Theories vs. Connectionism

Main Reading:

Further Reading:


WEEK 7: The Chinese Room Argument

Main Reading:

Further Reading:


WEEK 8: The Chinese Room Revisited

Main Reading:

Further Reading:


WEEK 9: The Limits of Formalisation

Main Reading:

Further Reading:


**WEEK 10: The Mathematical Mind**

**Main Reading:**
*(NB: Read all of the above in chronological order)*

**Further Reading:**

**WEEK 11: Embodied Cognition**

**Main Reading:**

**Further Reading:**

**WEEK 12: Consciousness**

**Main Reading:**

**Further Reading:**
WEEK 13: Representation

Main Reading:

Further Reading:

WEEK 14: Expert Systems

Main Reading:

Further Reading:

WEEK 15: The Frame Problem

Main Reading:

Further Reading:
WEEK 16: Superhuman Intelligence

Main Reading:
http://web.media.mit.edu/~minsky/papers/sciam.inherit.txt

Further Reading:
Moravec, H. (1994) ‘Robots Inherit Human Minds’, talk found at:
http://www.frc.ri.cmu.edu/~hpm/project.archive/general.articles/1995/RobotMind.talk.html