General Information:

Type of course: Compact Seminar
ECTS-Credits: 5 or 2.5 respectively
Language: English
Time: June, 1st 12:00 - 17:00; June 2nd 11:00 – 16:00; June 15th 12:00 – 17:00 and June 16th 11:00 - 16:00
Room: Stinzingstraße 12, room I/6 (June 1st and June 15th) and I/7 (June 2nd and June 16th)

Registration:

You can register via StudOn for the course entitled *"Philosophical Issues of Artificial Intelligence (Parallelgruppe 2)"*. Here is the link: https://www.studon.fau.de/studon/goto.php?target=crs_5165689
You can also find the course on Campo via this link:
https://www.campo.fau.de/qisserver/pages/startFlow.xhtml?_flowId=searchCourseNonStaffflow&_flowExecutionKey=e1s3

Description of the course:

Focusing on foundational texts and classical arguments from philosophy, cognitive science and computer science, the course offers an introduction to foundational philosophical issues of artificial intelligence. It will cover questions such as: What is intelligence and how do we know a machine is intelligent? How is the intelligence of artificial systems different from ours? Is the right kind of computational structure sufficient for the possession of a mind? Is the human mind a kind of computer? Could a sophisticated robot think, feel, or be consciously aware? Will there ever be super intelligent machines?

The course also introduces philosophy as a practical activity by developing student skills in reading, analysis, reasoning, and argumentation, and fostering appreciation for opposing points of view and of those who hold them.

Aims of the course:

This course is primarily concerned with teaching skills that are necessary in any scientific study: Close reading, precise analysis, and stringent argumentation. In the course, we will take apart arguments, identify their assumptions and implicit presuppositions, criticize them, defend them against criticism, and criticize again the defense. You will need to practice these skills. This means that besides becoming familiar with philosophically relevant questions and arguments concerning AI research, you will engage in reading and discussing philosophical texts. In the exercises for the proofs of readings, for example, you will be asked to explain various theories in broad terms and then to offer a critique or defense.

How we will proceed:

I assume that you will independently prepare and follow up the texts for the respective session and will usually not provide a summary of the content. However, for particularly difficult texts or text passages, I will help with explanations. You should therefore assume that we will use the time for the sessions intensively to analyze and criticize the theses and arguments presented. You can best participate in this process if you make use of the self-study phases and have already read the texts and made some notes before the respective session. Please take this point seriously.

Completion of the course:

Proofs of reading: I will pose a question for each of the texts assigned in schedule via the StudOn page starting on May 17th. You are expected to answer at least four of these questions

in written form by preparing a text of no more than 400 words for each question. Which assigned texts and questions you choose to write about is up to you, but make sure that you pick on from each of the four days of the seminar. You must upload your text on the StudOn page before the start of the seminar session on the day we will discuss the respective target paper. The exercises are designed to help you sort out your thoughts and get them down on paper in an orderly fashion.

End-term paper (5 ECTS option only): At the end of the course, you will explore a problem of your choice in greater detail by writing a term paper of ~5000 words (at 12pt, this is equivalent to about 10 pages). The problem explored should be directly related to one or more topics of the seminar. Please contact me early on, if you are interested in a topic so that we can determine a suitable question. Please send the term paper as an e-mail attachment in docx or odt format to jon.leefmann@fau.de by September 15, 2023. Please name your file according to the scheme [NAME]_[TITLE OF WORK].docx. For more information on the structure and writing of term papers, take a look on the StudOn page.

Readings:

In addition to the obligatory readings indicated in the schedule (mainly classical texts in AI, cognitive science, and philosophy of AI) it will be of help to take a look on the following sources for a general overview of the field:

- Bringsjord, S. and Govindarajulu, N.S. (2022) 'Artificial Intelligence', in Edward N. Zalta and Uri Nodelman (eds.) *The Stanford Encyclopedia of Philosophy,* 2022nd edn.: Metaphysics Research Lab, Stanford University. (https://plato.stanford.edu/entries/artificial-intelligence/)

- Carter, M. (2007) *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence*. Edinburgh: Edinburgh University Press.

- Frankish, K. and Ramsey, W.M. (eds.) (2014) *The Cambridge Handbook of Artificial Intelligence*: Cambridge University Press. (available online via the university library: https://www.cambridge.org/core/books/cambridge-handbook-of-artificial-intelligence/3DCB2E04739722A99EDE86B7A34A30E3)

Schedule:

Topic	Readings
Thursday, June 1ˢᵗ (12 am – 5 pm, room I/6, Stinzingstraße 12)	
Introduction	-
Intelligence and Behavior	Turing, A. M. (2000) 'Computing Machinery and Intelligence (1950)', in Haugeland, J. (ed) *Mind design II: Philosophy, psychology, artificial intelligence,* 3rd edn, Cambridge, Mass., MIT Press, pp. 29–56.
	Putnam, H. (2004) 'Brains and Behavior (1961)', in Heil, J. (ed.) *Philosophy of Mind: A guide and anthology.* Oxford: Oxford Univ. Press, pp. 96–104.
Friday, June 2ⁿᵈ (11 am – 4 pm, room I/7, Stinzingstraße 12)	
Computer Functionalism	Block, N. (1998) 'The Computer Model of the Mind', in Smith, E.E., Osherson, D.N. and Lasnik, H. (eds.) *An Invitation to Cognitive Science,* 2nd edn. (A Bradford book, 3). Cambridge, Mass.: MIT Press, pp. 377–425.
	Newell, A. and Simon, H. (2000) 'Computer Science as Empirical Inquiry. Symbols and Search (1976)', in Haugeland, J. (ed.) *Mind design II: Philosophy, psychology, artificial intelligence,* 3rd edn. Cambridge, Mass., MIT Press.

Thursday, June 15th (12 am – 5 pm, room I/6, Stinzingstraße 12)	
The Chinese Room Argument	Searle, J. R. (2000) 'Minds, Brains, and Programs (1980)', in Haugeland, J. (ed) *Mind design II: Philosophy, psychology, artificial intelligence,* 3rd edn. Cambridge, Mass., MIT Press, pp. 183–204.
	Boden, M. (2004) 'Escaping the Chinese Room (1988)', in Heil, J. (ed.) *Philosophy of Mind: A guide and anthology.* Oxford: Oxford Univ. Press, pp. 253–266.
The Frame Problem of AI	Dennett, D.C. (1988) 'Cognitive Wheels. The Frame Problem of AI (1984)', in Pylyshyn, Z.W. (ed.) *The Robot's Dilemma: The Frame Problem in Artificial Intelligence,* 2nd edn. Norwood, NJ: Ablex Publ, pp. 41–63.
	Miracchi, L. (2020) 'Updating the Frame Problem for AI Research', *Journal of Artificial Intelligence and Consciousness,* 07(02), pp. 217–230. doi: 10.1142/S2705078520500113
Friday, June 16th (11 am – 4 pm, room I/7, Stinzingstraße 12)	
Language Models and Connectionism	Sun, R. (2014) 'Connectionism and neural networks', in Frankish, K. and Ramsey, W.M. (eds.) *The Cambridge Handbook of Artificial Intelligence:* Cambridge University Press, pp. 108–127.
	Chalmers, D. (2022) 'Could a large language model be conscious?' Video: https://nips.cc/virtual/2022/invited-talk/55867
	Chomsky, N.; Roberts, I.; Watumull J. (2023) The false promise of ChatGPT, *The New York Times*, Mach 8th 2023.
Superintelligent Machines	Chalmers, D. (2010) 'The Singularity: A Philosophical Analysis', *Journal of Consciousness Studies*, 17(9-10), pp. 7–65. (Excerpt pp. 7-30)
Closing Session	-