Stockholm University
Department of Psychology

Course Syllabus
Artificial General Intelligence, Fall 2022
(7.5 ECTS/hp)

1. Type of Course
Course at the doctoral level, Department of Psychology, Stockholm University. The course is a collaboration between Stockholm University and the cross-disciplinary research centre Digital Futures. The course will be held physically in Stockholm, as well as via Zoom.

2. Course dates
All course dates are mandatory. Individual completion will be possible. Planned course dates are eight Thursdays 4pm-6pm, starting September 29th, ending November 24th.

3. Course Leader
Robert Johansson, robert.johansson@psychology.su.se

4. Teachers
Pei Wang, Guest Professor, pei.wang@psychology.su.se
Patrick Hammer, Postdoc, patrick.hammer@psychology.su.se
Robert Johansson, Associate Professor, robert.johansson@psychology.su.se

5. Course description
The original goal of the AI field was to make machines that could think like humans - that is, computer systems with human-like general-purpose intelligence. This has been a very difficult task, which has led a majority of AI researchers to focus on what has been called “narrow AI” - the construction of AI systems that solve specific tasks in narrow domains.

Importantly though, more researchers are recognizing the necessity of returning to the original goals of AI. During the last 10-15 years there has been an increasingly stronger call for a transition back to confronting the more difficult issues of “human level intelligence” and more broadly “artificial general intelligence” (AGI).

6. Course activities
The course consists of a theoretical part, and a project part.

The theoretical part includes, but is not limited, to the following:

- Definitions of AGI
- Limitations and objections to AGI
- Strategies and techniques for achieving AGI
- Ethics of AGI

The project part of the course will be to conduct a project related to AGI. Examples include:

- Implementation of a subset of a particular AGI model described during the course

- Robotics project, where parts of the software is based on an AGI software platform
- Experiment that evaluates the cognitive capabilities of an AGI system
- A philosophical text describing an important problem in AGI from a new perspective

All students will decide a topic for their practical project in discussion with the course organizers. Preferably, the topics are chosen with a connection to students' respective subject/research area. There will be an opportunity for individual supervision on the respective projects.

7. Learning objectives

During the course, students are encouraged to find their own answers to the following question:

- What is AGI, accurately specified?
- Is it possible to build the AGI as specified?
- If AGI is possible, what is the most plausible way to achieve it?
- Even if we know how to achieve AGI, should we really do it?

8. Examination

The examination consists of two parts:

- A written paper (5-10 pages) where the student provide their own answers (with arguments) to the above four questions
- Presentation/demo of the project at an open seminar at the Department of Psychology

9. Previous knowledge

Admitted to postgraduate studies in Psychology, Cognitive Science, Computer Science, or a related discipline. Master students with sufficient previous knowledge are also welcome.

10. Grades

Pass or Fail

11. Literature

Scientific papers: (*243 pages in total*)

Everitt, T., Lea, G., & Hutter, M. (2018). AGI safety literature review. *arXiv preprint arXiv:1805.01109*.

Li, X., Hammer, P., Wang, P., & Xie, H. (2018, August). Functionalist emotion model in nars. In *International Conference on Artificial General Intelligence* (pp. 119-129). Springer, Cham.

Wang, P. (2012). Theories of artificial intelligence - Meta-theoretical considerations. In *Theoretical foundations of artificial general intelligence (pp. 305-323)*. Atlantis Press, Paris.

Wang, P., Hahm, C., & Hammer, P. (2022). A Model of Unified Perception and Cognition. *Frontiers in Artificial Intelligence*.

Wang, P., Li, X., & Hammer, P. (2018). Self in NARS, an AGI System. *Frontiers in Robotics and AI*, 5, 20.

Wang, P. (2020). A constructive explanation of consciousness. *Journal of Artificial Intelligence and Consciousness*, 7(02), 257-275.

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, 10(2), 1-37.

Wang, P. (2020). Commentaries and Author's Response. *Journal of Artificial General Intelligence*, 11(2).

Books, optional reading:

Bach, J. (2009). *Principles of synthetic intelligence: Psi, an architecture of motivated cognition*. Oxford: Oxford University Press

Hutter, M. (2005). *Universal artificial intelligence - sequential decisions based on algorithmic probability*. Berlin: Springer

Hawkins, J. (2021). *A thousand brains: a new theory of intelligence*. New York: Basic Books.

Yates, J., Immergut, M. & Graves, J. (2017). *The mind illuminated: a complete meditation guide integrating buddhist wisdom and brain science for greater mindfulness*. New York: Touchstone.

Scientific papers, optional reading:

Baars, B. J., & Franklin, S. (2009). Consciousness is computational: The LIDA model of global workspace theory. *International Journal of Machine Consciousness, 1(01)*, 23-32.

Franklin, S., Madl, T., D'mello, S., & Snaider, J. (2013). LIDA: A systems-level architecture for cognition, emotion, and learning. *IEEE Transactions on Autonomous Mental Development, 6(1)*, 19-41.

Franklin, S., Strain, S., McCall, R., & Baars, B. (2013). Conceptual commitments of the LIDA model of cognition. *Journal of Artificial General Intelligence, 4(2)*, 1-22.

Kugele, S., & Franklin, S. (2021). Learning in LIDA. *Cognitive Systems Research, 66*, 176-200.

Lenat, D. B. (2022). The Rise and Fall and Rise of Expert Systems. *IEEE Annals of the History of Computing*.