

Philosophy of AI

Hauptseminar, SoSe 2024

Konstantin Genin and Hong Yu Wong

konstantin.genin@gmail.com / hong-yu.wong@uni-tuebingen.de

12-14; Thursdays, Room X, Alte Burse

1. April 18: Introduction - Turing Test/ Intentional stance

Turing, A.M. (1950) "Computing Machinery and Intelligence". *Mind* 49: 433-460.

Dennett, D. (1981) "True Believers" (in *Mind Design II*)

2. April 25: Chinese Room Argument/ Dreyfus

Searle, John R. (1980) "Minds, brains, and programs." *Behavioral and brain sciences* 3.3: 417-424.

Dreyfus, Hubert. (1979). "From Micro-Worlds to Knowledge Representation: AI at an Impasse", *Mind Design II: Philosophy, Psychology, and Artificial Intelligence*, ed. John Haugeland, MIT Press, pp. 143-182.

3. May 2: Deep Learning/ Comparative approaches to cognition

Buckner, Cameron. (2024) "What is Deep Learning, and How Should We Evaluate its Potential?" *From Deep Learning To Rational Machines*. Oxford University Press, pp. 48-93.

Halina, Marta (2021). Insightful artificial intelligence. *Mind & Language*, 36(2), 315-329.

4. June 6: Machine Learning and Kant (Richard Evans 'The Apperception Engine')

Evans, Richard. (2022) "The Apperception Engine". *Kant and Artificial Intelligence*, ed. by H. Kim and D. Schönecker, De Gruyter, pp. 39-104.

5. June 13: Reliability

Grote, Thomas, Genin, Konstantin and Emily Sullivan. (forthcoming) "Reliability in Machine Learning." *Philosophy Compass*.

Freiesleben, Timo, and Thomas Grote. (2023) "Beyond generalization: a theory of robustness in machine learning." *Synthese* 202.4: 109.

6. June 20: Fairness

Vredenburg, Kate. (2023) "[Fairness](#)", *Oxford Handbook on AI Governance*.

Green, Ben. (2022) "Escaping the impossibility of fairness: From formal to substantive algorithmic fairness." *Philosophy & Technology* 35.4: 90.

7. June 27: Justice

Kuppler, Matthias, et al. (2022) "From fair predictions to just decisions? Conceptualizing algorithmic fairness and distributive justice in the context of data-driven decision-making." *Frontiers in sociology* 7.

Beigang, Fabian. (2022) "On the advantages of distinguishing between predictive and allocative fairness in algorithmic decision-making." *Minds and Machines* 32.4: 655-682.

8. July 11: Explainability, Transparency, Interpretability, Trust

Creel, Kathleen A. (2020) "Transparency in complex computational systems." *Philosophy of Science* 87.4: 568-589.

Zerilli, John, et al. (2019) "Transparency in algorithmic and human decision-making: is there a double standard?." *Philosophy & Technology* 32: 661-683.

9. July 18: Presentations

Requirements for Credit

***Shared requirements for all degrees and credit combinations:**

- (a) Attendance (2 absences allowed without grounds)
- (b) 1 Short Presentation + 1 Project Presentation (July 18)
- (c) Complete all homework assignments (even if absent from class)

MSc ML / MSc Cogsci / MSc Computational Neuroscience

3 ECTS graded

* + 2000 Word essay on a topic to be determined in consultation with the instructors

BSc Cogsci

3 ECTS graded

* + 1500 Word essay on a topic to be determined in consultation with the instructors

MA Philosophy

3 ECTS ungraded

* + 1500 Word essay on a topic to be determined in consultation with the instructors

12 ECTS graded

* + 5000 Word essay on a topic to be determined in consultation with the instructors

MEd Philosophy/Ethics (2021 MHB)

3 ECTS ungraded

* + 1500 Word essay on a topic to be determined in consultation with the instructors

8 ECTS graded

* + 4000 Word essay on a topic to be determined in consultation with the instructors

MEd Philosophy/Ethics (2018 MHB)

3 ECTS ungraded + 5 ECTS graded

* + 4000 Word essay on a topic to be determined in consultation with the instructors

BA Philosophy (both 2021/2013 MHB) & BEd Philosophy/Ethics (2021 MHB)

6 ECTS graded

* + 3000 Word essay on a topic to be determined in consultation with the instructors

BEd Philosophy/Ethics (2015 MHB)

3 ECTS ungraded

* + 1500 Word essay on a topic to be determined in consultation with the instructors

3 ECTS ungraded + 6 ECTS graded

* + 4000 Word essay on a topic to be determined in consultation with the instructors

Other Tübingen Students

Please contact the instructors to make an arrangement. Tell us the number of ECTS you need for your degree program and whether it is graded.

Erasmus and Exchange Students

Please contact the instructors to make an arrangement. Tell us the number of ECTS you need for your university/degree and whether it is graded.

DEADLINE FOR ALL ESSAYS: 15.09.2024