EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

machine learning
new perspectives
for science

PhilML reading group

Venue: AI Research Building, **3rd-floor** seminar room, Maria-von-Linden-Straße 6, Tübingen.

The **PhilML** reading group discusses current topics in the philosophy of machine learning with a special focus on the philosophy of science. All interested students and researchers are welcome to join. Participants are asked to read the respective paper in advance.
Organized by Timo Freiesleben, Ben Höltgen, and Sebastian Zezulka.
More information on **talks.tue.ai.**

	Readings
16.04.24, 1-2 pm	**Variance in Fair Classification** Feder Cooper, A., et al (2024). Arbitrariness and Social Prediction: The Confounding Role of Variance in Fair Classification. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 20, pp. 22004-22012).
30.04.24, 1-2 pm	**Decision-Making without Counterfactuals** Dawid, A. P., & Senn, S. (2023). Personalised decision-making without counterfactuals. *arXiv preprint arXiv:2301.11976*.
14.05.24, 1-2 pm	**Understanding Deep Learning** Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, *64*(3), 107-115.
28.05.24, 1-2 pm	**Severe Testing and Methodological Underdetermination** Mayo, D. G. (1997). Severe tests, arguing from error, and methodological underdetermination. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *86*(3), 243-266.
11.06.24, 1-2 pm	**Measurement and Fairness** Jacobs, A. Z., & Wallach, H. (2021, March). Measurement and fairness. In *Proceedings of the 2021 ACM FAccT Conference* (pp. 375-385).
25.06.24, 1-2 pm	**Randomized Control Trials** Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social science & Medicine*, *210*, 2-21.
09.07.24, 1-2 pm	**Emergent Abilities in LLMs** Schaeffer, R., Miranda, B., & Koyejo, S. (2024). Are emergent abilities of large language models a mirage?. *Advances in Neural Information Processing Systems*, *36*.