



## **Philosophische Aspekte des maschinellen Lernens**

**Seminar im Wintersemester 2021/2022**

### **ORGANISATORISCHES**

Zeit:	Dienstag, 9:45-11:15 Uhr
Ort:	M 11.91
Modulzuordnung:	Theoretische Philosophie I/II (HF: KM5/6, BE: BE10) Interdisziplinäre Themen I/II (HF: KM9/10, NF: KM3, BE: BE12)
Dozent:	Oliver Buchholz
E-Mail:	<a href="mailto:oliver.buchholz@uni-tuebingen.de">oliver.buchholz@uni-tuebingen.de</a>
Sprechstunde:	nach Vereinbarung

### **KURZBESCHREIBUNG**

Maschinelles Lernen (ML) ist innerhalb weniger Jahre zu einer nahezu allgegenwärtigen Technologie geworden, die einerseits große Chancen, andererseits jedoch auch beträchtliche Risiken birgt. Das Spannungsfeld von Chancen und Risiken des ML stellt die Gesellschaft vor fundamentale Fragen – und ist nicht zuletzt deshalb philosophisch relevant.

Ziel dieses Seminars ist es, einen Überblick über einige der wichtigsten philosophischen Aspekte des ML zu geben. Nach einer Einführung in die grundlegende Funktionsweise des ML soll zunächst die erkenntnistheoretische Frage nach den Folgen der Intransparenz komplexer ML-Algorithmen diskutiert werden: Können wir den Ergebnissen eines ML-Algorithmus vertrauen, obwohl ihr Zustandekommen nicht exakt nachvollziehbar ist? Insbesondere vor diesem Hintergrund wird anschließend aus einer wissenschaftstheoretischen Perspektive zu überprüfen sein, ob sich der Einsatz von ML als Werkzeug in der Wissenschaft rechtfertigen lässt. Darüber hinaus werden ethische Probleme, die sich aus der (Un)fairness von ML-Algorithmen ergeben, Gegenstand des Seminars sein, bevor mithilfe der politischen Theorie Diskriminierungspotentiale des ML gegenüber bestimmten gesellschaftlichen Gruppen analysiert werden. Im Laufe des Seminars wird sich somit zeigen, dass das ML nicht nur unseren Alltag, sondern darüber hinaus nahezu alle Teilbereiche der Philosophie berührt.



## **LERNZIELE**

Nach einem erfolgreichen Besuch des Seminars

- verfügen Sie über ein Verständnis der grundlegenden Funktionsweise des ML
- kennen Sie die zentralen Debatten in der Philosophie des ML
- können Sie ein zum Themenbereich gehöriges Problem mündlich und schriftlich ausarbeiten und erörtern.

## **STUDIEN- UND PRÜFUNGSLEISTUNG**

Die *Prüfungsvorleistung* in diesem Seminar umfasst:

- Gründliche Vorbereitung der Lektüre für jede Sitzung
- Anwesenheit und aktive Beteiligung im Seminar
- Einen kurzen, unbenoteten Impulsvortrag

Die *Prüfungsleistung* besteht in einer benoteten Hausarbeit mit 10-15 Seiten (Abgabe: 31.03.2022). Die Gesamtnote für das Seminar ist identisch mit der Note für die Hausarbeit. Für Module, die eine unbenotete Studienleistung vorsehen, ist lediglich die Prüfungsvorleistung zu erbringen.

## **SONSTIGE HINWEISE**

Bitte treten Sie der ILIAS-Gruppe für das Seminar bei. Dort finden Sie neben Informationen zum Seminar auch die Literatur für alle Sitzungen.



## SITZUNGEN

- 19.10. Einführung
- 26.10. Was ist ML? Methodische Grundlagen.  
Literatur: v. Luxburg (2020). Wie funktioniert maschinelles Lernen? Unveröffentlichtes Manuskript.  
Zur Vertiefung: Jordan, Mitchell (2015). Machine Learning: Trends, Perspectives, and Prospects. *Science*, 329(6245), 255-260.
- 02.11. Warum ML? Philosophische Relevanz (I).  
Literatur: Thagard (1990). Philosophy and Machine Learning. *Canadian Journal of Philosophy*, 20(2), 261-276.
- 09.11. Was ist das Besondere an neuronalen Netzen? Philosophische Relevanz (II).  
Literatur: Buckner (2019). Deep Learning: A Philosophical Introduction. *Philosophy Compass*, 14(10), e12625.  
Zur Vertiefung: LeCun, Bengio, Hinton (2015). Deep Learning. *Nature*, 521(7553), 436-444.
- 16.11. Warum ist ML oft intransparent? Erkenntnistheoretische Probleme (I).  
Literatur: Schubbach (2021). Judging Machines: Philosophical Aspects of Deep Learning. *Synthese*, 198(2), 1807-1827.
- 23.11. Wie kann die Intransparenz des ML überwunden werden? Erkenntnistheoretische Probleme (II).  
Literatur: Lipton (2018). The Mythos of Model Interpretability. *Queue*, 16(3), 31-57.
- 30.11. Sollte die Intransparenz des ML überwunden werden? Erkenntnistheoretische Probleme (III).  
Literatur: Krishnan (2020). Against Interpretability: A Critical Examination of the Interpretability Problem in Machine Learning. *Philosophy & Technology*, 33(3), 487-502.
- 07.12. Eignet sich ML als Werkzeug in der Wissenschaft? Wissenschaftstheoretische Betrachtungen (I).  
Literatur: Williamson (2010). The Philosophy of Science and its Relation to Machine Learning. In Gaber (ed.), *Scientific Data Mining and Knowledge Discovery*. Springer, Berlin, 77-89.
- 14.12. Kann ML wissenschaftliches Verständnis fördern? Wissenschaftstheoretische Betrachtungen (II).  
Literatur: Sullivan (2020). Understanding from Machine Learning Models. *The British Journal for the Philosophy of Science*, forthcoming.



- 21.12. Repräsentiert ML empirische Phänomene? Wissenschaftstheoretische Betrachtungen (III).  
Literatur: Knüsel, Baumberger (2020). Understanding Climate Phenomena with Data-driven Models. *Studies in History and Philosophy of Science Part A*, 84, 46-56.
- 11.01. Präzise Vorhersage oder unmoralische Generalisierung? Ethische Probleme (I).  
Literatur: Munton (2019). Beyond Accuracy: Epistemic Flaws with Statistical Generalizations. *Philosophical Issues*, 29(1), 228-240.
- 18.01. Warum kann ML unfair sein? Ethische Probleme (II).  
Literatur: Mittelstadt, et al. (2016). The Ethics of Algorithms: Mapping the Debate. *Big Data & Society*, 3(2), 1-21.
- 25.01. Wie kann die Unfairness des ML überwunden werden? Ethische Probleme (III).  
Literatur: Binns (2018). Fairness in Machine Learning: Lessons from Political Philosophy. *Proceedings of the Conference on Fairness, Accountability and Transparency*, 1-11.
- 01.02. Ist ML kolonialistisch? Eine postkoloniale Perspektive.  
Literatur: Mohamed, Png, Isaac (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, 33(4), 659-684.
- 08.02. Rück- und Ausblick