Epistemological Issues of
Machine Learning
in Science

Date: 27.-28.02.2024

Place: Chaudoire Pavillon, TU Dortmund

UDNN

tu dortmund university

LAMARR INSTITUTE FOR MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Emmy Noether-Programm DFG Deutsche Forschungsgemeinschaft

This workshop is an event co-organized by the Emmy Noether Group

UDNN: Scientific Understanding and Deep Neural Networks

and the

Lamarr Institute for Machine Learning and Artificial Intelligence

https://udnn.tu-dortmund.de/ https://lamarr-institute.org/

Main organizers: Annika N. Schuster, Frauke Stoll, Leon Augustin & Florian J. Boge

Contents

About

Epistemological Issues of Machine Learning in Science

With impressive advances in Machine Learning (ML) and particularly Deep Learning, Artificial Intelligence is currently taking science by storm. This workshop brings together top scientists and philosophers working on fundamental issues connected to the use of Machine Learning in science. The workshop marks the launch of the DFG-funded Emmy Noether Group UDNN: Scientific Understanding and Deep Neural Networks, and is co-organized with the Lamarr Institute for Machine Learning and Artificial Intelligence and co-funded by the Department for Humanities and Theology at TU Dortmund University. Topics include, but are not restricted to:

1. The relation between prediction and discovery on the one hand, and explanation and understanding on the other, in fields of science that heavily rely on ML methods

2. The key issues in identifying genuine discoveries and stable predictions by ML systems

3. Core conceptions of “explanation” involved in the field of eXplainable AI (XAI), and their relation to philosophical theories of understanding and explanation

4. Present limitations associated with ML’s predictive power and what may be needed to overcome them

5. The connection between ML and traditional scientific means for prediction and discovery, such as theories, models, and experiments

6. Our present understanding of ML itself and its limitations

Main Organizers: Annika N. Schuster, Frauke Stoll, Leon Augustin & Florian J. Boge

Timetable

Tuesday, 27 of February

9:00–9:15	**Arrival and Coffee**		
9:15–9:20	**Opening Words**		
9:20–10:05		**Henk de Regt** Radboud University, Nijmegen	Can machines acquire scientific understanding?
10:05–10:50		**David Watson** King‘s College London	Richness revisited: Clustering and PAC learnability
10:50–11:05	**Coffee**		
11:05–11:50		**Tom Sterkenburg** LMU Munich	Occam’s razor in machine learning
11:50–12:20		**Annika Schuster** TU Dortmund	A new pathway: From objectual to explanatory understanding with AlphaFold2
12:20–13:20	**Lunch**		
13:20–14:05		**Dominik Elsässer** TU Dortmund	Is knowledge forever? An astronomical perspective
14:05–14:50		**Wolfgang Rhode** TU Dortmund	ML-driven knowledge gain in physics
14:50–15:35		**Brigitte Falkenburg** TU Dortmund	Data, theories, and machine learning in astroparticle physics
15:35–15:50	**Coffee**		
15:50–16:35		**Lena Kästner** University of Bayreuth	Opacity as a stepping stone
16:35–17:20		**Jürgen Bajorath** University of Bonn	Explainable machine learning in drug discovery
18:30–19:30	**Dortmunder U**		
20:00	**Conference Dinner**		

Wednesday, 28 of February

9:00–9:15	**Arrival and Coffee**		
9:15–10:00		**Konstantin Genin** University of Tübingen	From the fair distribution of predictions to the fair distribution of social goods: Evaluating the impact of fair machine learning on long-term unemployment
10:00–10:45		**Marie-Jeanne Lesot** Sorbonne	Explainable AI and trustworthy AI: A relation to discuss
10:45–11:00	**Coffee**		
11:00–11:45		**Mario Krenn** MPI Erlangen	Towards an artificial muse for new ideas in science
11:45–12:15		**Frauke Stoll** TU Dortmund	Navigating the black box: Understanding particle physics with deep neural networks and eXplainable artifical intelligence
12:15–13:15	**Lunch**		
13:15–14:00		**Miriam Klopotek** University of Stuttgart	What can we learn from and through machine learning if the physics of many-body systems is behind it?
14:00–14:45		**Eva Schmidt** TU Dortmund	Stakes and understanding the decisions of artificial intelligent systems
14:15–15:30		**Axel Mosig** Ruhr University Bochum	A hypothesis-centric perspective on machine learning in biomedicine
15:30–15:45	**Coffee**		
15:45–16:30		**Kathleen Creel** Northeastern University	Concepts on the move: transparency, proxies, and conceptual engineering
16:30–17:15		**Florian Boge; Michael Krämer; Christian Zeitnitz** TU Dortmund; RWTH Aachen; BU Wuppertal	Deep learning for scientific discovery and the theory-freedom-robustness trade-off
17:15–17:20	**Closing Words**		

Abstracts

Tuesday 27th

Can machines acquire scientific understanding?

Henk de Regt

Radboud University, Nijmegen

In many areas of present-day science machine learning plays an increasingly important role. While it is clear that machines can assist scientists in their quest for understanding the world, a more controversial question is whether they can also generate scientific understanding independently of human scientists. The answer to this question obviously depends on how we define (scientific) understanding. In my talk I will outline my contextual theory of scientific understanding and explore the prospects of 'artificial scientific understanding' from the perspective of this theory.

Richness revisited: Clustering and PAC learnability

David Watson

King‘s College London

Clustering is ubiquitous in data science, but the theory behind it remains poorly developed. In a celebrated paper, Kleinberg (2002) proves that no clustering algorithm can simultaneously satisfy three seemingly unobjectionable axioms. The vast majority of replies to Kleinberg's impossibility theorem target the so-called "consistency“ axiom, which states (roughly) that a clustering algorithm's results should not change if we replace one distance measure with another that shrinks within-cluster distances and expands between-cluster distances. I take a different tack, instead targeting the "richness" axiom, which holds that a clustering algorithm should be able to learn any possible partition of the data. This, I argue, is an impossibly high bar that effectively demands more of an unsupervised learning algorithm than we can theoretically expect of any supervised classifier. Replacing the richness axiom with a PAC learnability criterion saves clustering from Kleinberg's impossibility result without any modifications to the consistency axiom. It also illustrates one possible strategy for unifying epistemological approaches across different machine learning paradigms.

Occam’s razor in machine learning

Tom Sterkenburg

LMU Munich

The principle of Occam’s razor tells us to seek simplicity in inductive inference. Occam’s razor makes a regular appearance in machine learning, and it is indeed frequently suggested that the mathematical theory of machine learning can offer us a justification for the principle: that is, an argument why a simplicity preference leads to better learning. At the same time, there also exist arguments, both in the philosophy and in the computer science literature, against such a justification.
In my talk, I will attempt to unite these opposing views by drawing out what kind of qualified justification for Occam’s razor may be had from statistical learning theory, the standard theoretical framework for machine learning. I will arrive at (with a label that is suggestive of its highly qualified nature) a "model-relative means-ends" justification, which, roughly, says that it is a good idea to try to model your assumptions in a maximally simple model class, to enjoy a maximally strong guarantee relative to this class. Finally, I will say something about the current-day discussion regarding the "generalization paradox of deep learning," that suggests the need for a whole new theoretical approach in machine learning, including a different guise of Occam’s razor.

A new pathway: From objectual to explanatory understanding with AlphaFold2

Annika Schuster

TU Dortmund

DeepMind’s AlphaFold2 (AF2) deep neural network (DNN) (Jumper et al. 2021) gained a lot of attention when it surpassed other algorithmic devices for protein structure predictions from amino acid sequences considerably in the last Critical Assessment of protein Structure Prediction. Critical voices, however, remarked that the most important questions concerning protein folding are still unanswered. Due to the high dimensionality of the data they process and of the network architecture there is no straightforward way of understanding which features of the input data were responsible for their success. Objectual and explanatory understanding as two types of understanding commonly distinguished in the literature are of particular interest with regards to DNNs. Building on the case study of AF2 in protein biology, I will argue that the relationship of DNNs to science, explanation and understanding is best described as a two-step adaptive process. In detail, building on how scientists actually work with AF2 predictions, I will show that, in the first instance, DNNs like AF2 increase objectual understanding. However, in a second step, this increase can, and often does, lead to additional explanatory understanding.

Is knowledge forever? An astronomical perspective

Dominik Elsässer

TU Dortmund

Astronomy is one of the oldest cultural and scientific endeavors humankind engages in. For millennia, the collection and dissemination of knowledge about the Universe was based on low-volume data from the narrow sliver of the visual spectrum out of all electromagnetic radiation, processed by the human intellect and resulting in low-volume datasets preserved in oral tradition or written form. The goal was to thus preserve all observations once made in an ideal case for indefinite time, for posterity. This has changed dramatically. Today, we collect data not only across all wavebands in the electromagnetic spectrum, but also use completely different information carriers like neutrinos, charged particles, and gravitational waves. The complexity of the datasets has risen to a point that precludes procession by the human intellect alone, and in many cases the sheer volume of data rules out indefinite preservation of raw data. In this talk, I aim to discuss epistemological challenges arising from the need of taking a human decision on which subsets of raw data, which meta-data, and which analysis packages to preserve under the above mentioned harsh constraints, and under the unique astronomical border condition of growing reliance on machine learning methods in “experiments”that in some cases can by their very nature never be repeated. I will try to outline strategies employed in present leading astroparticle experiments, and discuss the need for the prioritization of formats that constrain as little as possible the intellectual freedom of future generations to use our scientific heritage for testing theories we may not even know of today.

ML-driven knowledge gain in physics

Wolfgang Rhode

TU Dortmund

All that can be experimentally measured are the electrical charges, from which the time and location intervals they are counted in, are known. These charges are always a consequence of all the physical processes outside the detector that correlate to the measurement. The statistical theories of thermodynamics and quantum electrodynamics always play a role here because every setup is subject to at least their effects. The number of such individual location- and time-dependent charge counts recorded in an experiment can be vast. Thousands of dimensions of experimental data are not an oddity. Such amounts of data can only be analyzed using machine learning methods. To our good fortune, all the available physics knowledge can be translated into a “virtual reality“ via simulations, which is used to optimize those machine learning algorithms and determine how well results can be reproduced and how significant their uncertainties are. With such well-understood machine learning algorithms, individual, meaningful events can be selected from an extensive data set in a desired manner. These events can then be used to solve the Inverse Problem associated with the measurement, i.e., the question of the conditions under which the cause can be deduced from the effect. Once the Inverse Problem has been solved, when instead of charges, locations, and times, measurement points in physical units and their uncertainties, are given, comparisons can be made between rationalistically devised theories and the measurements. The talk will discuss this ML-driven approach to gaining knowledge due to the necessary probabilistic character of every measurement called “probabilistic rationalism“.

Data, theories, and machine learning in astroparticle physics

Brigitte Falkenburg

TU Dortmund

After recapitulating the theory-data relation and the meaning of probability in physics, I discuss the role of computer simulations and ML in the measurements of astroparticle physics. In the probabilistic analysis of the big data processed here, computer simulations enter an iterative process of data correction that is optimized by Machine Learning (ML). In contrast to the experiments of particle physics at accelerators, the data processing of astroparticle physics no longer aims at the reconstruction of individual particle tracks but directly at the probability distributions measured by the particle detectors. Hence the measured data obtained from the raw data by the ML data processing procedures are intrinsically probabilistic. I will discuss this change of what is called "the data" under three aspects: (1) the opacity of ML; (2) the theory-ladenness of the data; and (3) the ontological question of what is measured, i.e., the kind of phenomena underlying the data of astroparticle physics.

References

[1] B.Falkenburg, Computer Simulation in Data Analysis: A Case Study from Particle Physics. Revised version, submitted to: Studies in History and Philosophy of Modern Physics.

[2] B. Falkenburg, Teilchen und Wellen als kosmische Boten: Eine philosophische Analyse. In: Oliver Passon et al., Kohärenz im Unterricht der Elementarteilchenphysik. Berlin: Springer Spektrum 2020, 37-54.

[3] B. Falkenburg, On the Contributions of Astroparticle Physics to Cosmology. In: Studies in the History and Philosophy of Modern Physics46 (2014,) 97–108.

[4] B. Falkenburg, Pragmatic Unification, Observation and Realism in Astroparticle Physics. In: General Journal for Phil. of Science 43 (2012), 327-345.

[5] B. Falkenburg, From Waves to Particles and Quantum Probabilities. In: B. Falkenburg & W. Rhode (eds.), From Ultrarays to Astroparticles. A Historical Introduction to Astroparticle Physics. Springer: Dordrecht 2012, 265-295.http://www.springer.com/gp/book/9789400754218

[6] B. Falkenburg, Particle Metaphysics. A Critical Account of Subatomic Reality. Heidelberg: Springer 2007. http://www.springer.com/gp/book/9783540337317

Opacity as a stepping stone

Lena Kästner

University of Bayreuth

Modern AI systems are often based on powerful machine-learning (ML) techniques; as a result, they can often make highly accurate predictions but are usually hardly intelligible, even to their developers. This opacity presents a major challenge for deployers, regulators and users of alike, specifically with respect to questions of safety, reliability and trustworthiness of AI systems. This paper discusses how AI systems' opacity might best be addressed. I shall argue that addressing opacity requires the collaboration of different domain experts engaging in coordinated epistemic activities and operations. By applying systematic discovery strategies familiar from the life sciences, domain experts can work towards uncovering the systems' overall functional architecture and thus render opaque systems mechanistically interpretable. Once a system is mechanistically interpretable, it may be re-engineered to build a transparent system that is arguably safer, does not utilize unfair associations and is less likely to suffer from unexpected failures. Thus, despite all the issues that black box systems raise, they can be a stepping stone for the development of transparent ones.

Explainable machine learning in drug discovery

Jürgen Bajorath

University of Bonn, University of Washington

In pharmaceutical research, explanatory methods for machine learning (ML), a part of explainable artificial intelligence (XAI), gain in importance as the complexity of ML models and predictions increases. In interdisciplinary environments, rationalizing predictions is of high relevance for increasing the acceptance of ML to aid in experimental design. The application of feature attribution methods for quantifying feature importance is often combined with visualization techniques to provide a biologically or chemically intuitive access to predictions. An XAI approach adapted for exploring structure-activity relationships is discussed. As an exemplary application, predictions of multi-target compounds are analyzed, also emphasizing the fundamental difference between correlation and causality in ML. The introduced methodological framework is applicable to explain molecular ML models and, if causality can be established, identify structural features that distinguish compounds with different properties.

Wednesday 28th

From the fair distribution of predictions to the fair distribution of social goods: Evaluating the impact of fair machine learning on long-term unemployment

Konstantin Genin

joint work with Sebastian Zezulka (University of Tübingen)

University of Tübingen

Algorithmic fairness focuses on the distribution of *predictions* at the time of *training*, rather than the distribution of *social goods* that arises after *deploying* the algorithm in a concrete social context. However, requiring a 'fair' distribution of predictions may undermine efforts at establishing a fair distribution of social goods. Our first contribution is conceptual: we argue that addressing the fundamental questions that motivates algorithmic fairness requires a notion of *prospective* fairness that anticipates the change in the distribution of social goods after deployment. Our second contribution is theoretical: we provide conditions under which this change is identified from pre-deployment data. That requires distinguishing between, and accounting for, different kinds of performative effects. In particular, we focus on the way predictions change policy decisions and, therefore, the distribution of social goods. Throughout, we are guided by an application from public administration: the use of algorithms to (1) predict who among the recently unemployed will remain unemployed in the long term and (2) target them with labor market programs. Our final contribution is empirical: using administrative data from the Swiss public employment service, we simulate how such policies would affect gender inequalities in long-term unemployment. When risk predictions are required to be 'fair', targeting decisions are less effective, undermining efforts to lower overall levels of long-term unemployment and to close the gender gap in long-term unemployment.

Explainable AI and trustworthy AI: A relation to discuss

Marie-Jeanne Lesot

Sorbonne Université

The proliferation of AI applications, in particular in sensitive domains, raises the crucial question of the trust users should put in models trained automatically. One of the directions proposed for establishing this trust is explored in the field of eXplainable Artifical Intelligence (XAI), in which models provide explanations or justifications for the decisions they make. The very definitions of trust and explanation remain a matter of debate, and the relationships between them are actually not as obvious as they are sometimes considered. This presentation proposes to discuss some questions that call for caution, a necessary preliminary step to the implementation of good practice in the XAI research field, which are themselves indispensable in the aim of establishing trust.

Towards an artificial muse for new ideas in science

Mario Krenn

Max Planck Institut Erlangen

Artificial intelligence (AI) is a potentially disruptive tool for physics and science in general. One crucial question is how this technology can contribute at a conceptual level to help acquire new scientific understanding or inspire new surprising ideas. I will talk about how AI can be used as an artificial muse in physics, which suggests surprising and unconventional ideas and techniques that the human scientist can interpret, understand and generalize to its fullest potential.

References

[1] Krenn, Kottmann, Tischler, Aspuru-Guzik, Conceptual understanding through efficient automated design of quantum optical experiments. Physical Review X 11(3), 031044 (2021).
[2] Krenn, Pollice, Guo, Aldeghi, Cervera-Lierta, Friederich, Gomes, Häse, Jinich, Nigam, Yao, Aspuru-Guzik, On scientific understanding with artificial intelligence. Nature Reviews Physics 4, 761–769 (2022).
[3] Krenn, Zeilinger, Predicting research trends with semantic and neural networks with an application in quantum physics. PNAS 117(4), 1910-1916 (2020).
[4] Krenn et al., Forecasting the future of artificial intelligence with machine learning-based link prediction in an exponentially growing knowledge network, Nature Machine Intelligence 5, 1326 (2023).

Navigating the black box: Understanding particle physics with deep neural networks and eXplainable artificial intelligence

Frauke Stoll

TU Dortmund

Building up on the arguments that gaining understanding from Deep Neural Networks (DNNs) requires a deeper insight into their inner workings, I will argue that the bridge to understanding in Particle Physics with DNNs can be found in the employment of techniques from eXplainable Artifical Intelligence (XAI). I will analyze these methods and show that there is a distinction between methods that target the DNN‘s functioning and those that target the underlying subject matter. Based on this distinction, I will argue that, in order to regain understanding in Particle Physics, in domains wherein the only successful prediction come from Machine Learning (ML), researchers will have to rely on XAI methods that decidedly target the subject matter rather than the ML model itself. More specifically, I will show how these XAI methods – in analogy to phenomenological models in Particle Physics - allow visualizations and qualitative predictions, thus paving the way towards intelligible physics models.

What can we learn from and through machine learning if the physics of many-body systems is behind it?

Miriam Klopotek

University of Stuttgart

In physics, the study of many-body systems hones in on how complex yet adaptive and even coordinated behavior emerges out of the motion of individual particles or agents. I am interested in how such physical emergence happens in artificial learning systems by finding specific analogies to many-body behavior. This has led me to develop several ‘physics-explainability’ techniques for ML that offer deepened insight into how algorithms work and what their limitations mean. At its heart, the problem of reducing the complexity of data may be akin to (time-dependent) coarse-graining in many-body systems. Moreover, some learning phenomena could be viewed as phase transformations. In a further step, I discuss how ML modeling – made intelligible in this way – could latch onto our cognition and lead to new insight in the realm of natural science. In fact, the analogy between ML and many-body dynamics is exact when information processing arises through physical dynamics – I discuss some results on reservoir computing through the non-equilibrium dynamics of swarm models.

Stakes and understanding the decisions of artificial intelligent systems

Eva Schmidt

TU Dortmund

Explainable artificial intelligence (XAI) aims to overcome the opacity of black box systems, i.e., to make them understandable to suitable stakeholders. In this paper, I investigate how understanding depends on how much is at stake in a context. I will support the intuition that understanding depends on the stakes with a pair of cases. I will further use this pair of cases to spell out how exactly the stakes affect understanding, particularly, understanding why. To do so, I will connect discussions of the concept of understanding with debates on pragmatic encroachment and on inductive risk. My aim, then, is to provide a pragmatic encroachment/inductive risk based account of how the stakes affect the understanding of the recipients of XAI explanations.

A hypothesis-centric perspective on machine learning in biomedicine

Axel Mosig

Ruhr University Bochum

Explainability has been discussed as an important cornerstone of transparent and trustworthy AI [1] especially in high-stakes fields like medicine. Considering the history of modern medicine beyond artificial intelligence, it is striking that transparency and trustworthiness are often obtained from scientific explanation, which relies on experimentally testable hypotheses. Approaches investigated in the field of eXplainable Artificial Intelligence (XAI), however, have a tendency to be rather expert-centric in the sense that many explainability methods are post-hoc and associate an interpretable representation to a given Machine Learning (ML) output that must then be assessed by a human expert. In recent work [2], we have introduced the framework of Falsifiable eXplanations for Artificial Intelligence (FXAI), wherein explanations are required to correspond to falsifiable hypotheses in the sense of empirical science. FXAI thus addresses the missing link between data-centric machine learning and the deductive aspects of the scientific method. We here elaborate on central aspects of FXAI, with a focus on medical applications, and address their implications for the trustworthiness of ML-systems. In particular, we argue that some core problems of XAI arise from how ML and its outcomes relate – or fail to relate – to evidence about physical reality. We present applications of FXAI in the identification of patterns of disease related to colorectal carcinoma. Finally, we consider the impact of experimental testability of explanations on trustworthiness.

References

[1] M. Ghassemi, L. Oakden-Rayner, and A. L. Beam. The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health, 3(11):e745–e750, 2021.
[2] D. Schuhmacher, S. Schörner, C. Küpper, F. Großerueschkamp, C. Sternemann, C. Lugnier, A.-L. Kraeft, H. Jütte, A. Tannapfel, A. Reinacher-Schick, et al. A framework for falsifiable explanations of machine learning models with an application in computational pathology. Medical Image Analysis, 82:102594, 2022.

Concepts on the move: transparency, proxies, and conceptual engineering

Kathleen Creel

Northeastern University

Intentionally side-stepping questions of whether machine learning systems can learn concepts tout court, Florian Boge (2024) introduces functional concept proxies (FCPs) as a deflationary category. If for a set of tasks and in a context, x fulfills the same causal roles as would a concept, x can be considered a functional proxy for that concept (p 11, 2024). In this paper, I connect functional concept proxies (and other deflationary interpretations) with the literature on conceptual engineering to ask whether complex machine learning systems such as large language models implicitly engineer scientific concepts or whether they allow concepts to wander significantly (Wilson 2008). In other words, for what do functional proxies stand?

Deep learning for scientific discovery and the theory-freedom-robustness trade-off

Florian Boge, Michael Krämer, and Christian Zeitnitz

TU Dortmund, RWTH Aachen, and BU Wuppertal

Machine Learning (ML) techniques such as Deep Neural Networks (DNNs) are of great promise in fueling scientific discovery. In High Energy Physics, they are expected to aid the detection of anomalies that are otherwise hard to find, thus promising novel discoveries without reliance on any specific theory or model and beyond what is currently humanly possible. But DNNs are also known for astonishing shortcomings, as they are vulnerable to 'adversarial examples'; data instances that are easily classifiable for humans but totally misclassified by DNNs. This raises questions of robustness. In our talk, we first argue that adversarial vulnerability is a double-edged sword: On the one hand, it shows that discerning DNNs' credible outputs from flukes requires some skill. On the other hand, adversarials exhibit DNNs' sensitivity to subtle, often human-inscrutable features that could also be scientifically productive (Buckner [2020]) – which are actually being utilized in anomaly detection. Hence, a comprehensive notion of performance-robustness is needed, which DNNs need to satisfy in order to be able to deliver genuine discoveries. Coining such a notion, we then offer a cautionary tale about DNNs' present utility for scientific discovery. As we shall argue, the achievement of performance-robustness implies limitations for the theory-freedom of ML-driven discovery.

Useful Information

General

Talks will be held at the **Rudolf-Chaudoire Pavillon** at TU Dortmund. It is situated on the **Campus Süd**.

Coffee breaks will be offered at the workshop location; **lunch** will take place at the nearby Archeteria.

Wi-Fi will be available during the conference via eduroam.

All speakers are invited to join us in visiting the exhibition “2hoch$x$ – Physik und Kunst zwischen Raum und Zeit” at the **Dortmunder U** after the talks on the first workshop day, which is located at Leonie-Reygers-Terrasse, 44137 Dortmund. We will guide you there from the hotel or from Dortmund main station.

The **conference dinner** will be held at the “Schönes Leben”, at Liebigstraße 23, 44139 Dortmund.

How to get around?

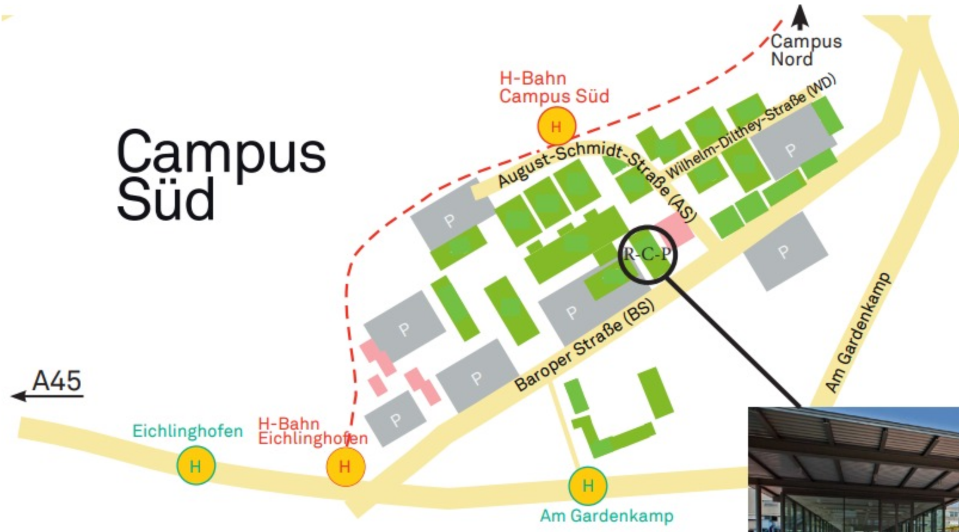
The **Campus Süd** can be reached via the **H-Bahn** (station: “Campus Süd”). If you travel with the S-Bahn, you arrive at the station “Dortmund Universität”. You can then find the H-Bahn just above the S-Bahn station. From the H-Bahn station Campus Süd, you‘ll need 5 minutes to the Rudolf-Chaudoire Pavillon.

- **S-Bahn:** line S1, station Dortmund Universität
- **H-Bahn:** station Campus Süd

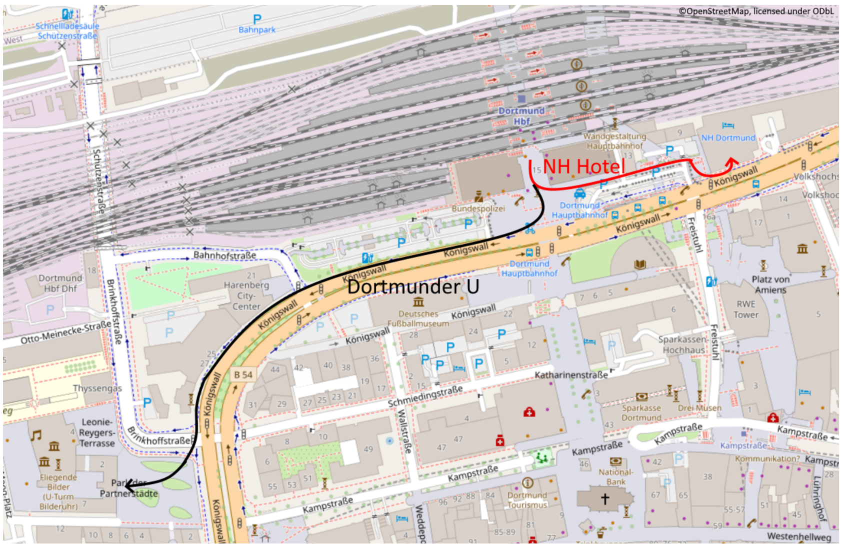
The **NH Hotel** is in the immediate proximity of Dortmund main station. Walk straight to the left when exiting the main station via the front exit; the hotel is within five minutes of walking distance.

The **Dortmunder U** is also in the proximity of Dortmund main station. You can reach it within ten minutes when exiting the main station to the right via the front exit. You can either consult the map on the next page or use your phone to navigate, but we recommend that you go there with the whole group.

The **Schönes Leben** is a 20 minutes walk from the Dortmunder U and about a 30 Minutes walk from the main station or the hotel. You can also reach it quickly from the train station “Möllerbrücke” or the subway station “Saarlandstraße”. Please make sure to keep your tickets when you choose to use public transportation.



Rudolf-Chaudoire-Pavillon
Baroper Straße 297, Campus Süd



Organization and Funding

This workshop is an event co-organized by the Emmy Noether Group *UDNN: Scientific Understanding and Deep Neural Networks* (DFG Grant 508844757) and the *Lamarr Institute for Machine Learning and Artificial Intelligence*, which is institutionally funded by the Federal Ministry of Education and Research (BMBF) and the state of North Rhine-Westphalia. The event is also generously supported by the *Department for Humanities and Theology* at TU Dortmund University.