Artificial Intelligence: Probabilistic Concepts

Spring school, March 16–20, 2026, Technical University of Darmstadt

Mini Courses:
Marco Mondelli
Andrea Montanari

Invited Speakers:
Jean Barbier
Lénaïc Chizat
Leif Döring

Organization:
Frank Aurzada
Volker Betz
Matthias Meiners

Contents

1 General Information

1.1 Registration

On Monday morning, starting from 8:00, registration is possible in the lobby of the lecture hall.

1.2 Lecture Hall

Location: Technische Universität Darmstadt. The registration and all lectures will take place in building S2|04, Hochschulstraße 8, 64289 Darmstadt in lecture hall S2|04 213. In the lecture hall, there are 2 large blackboards and 2 small blackboards and a projector.

1.3 Map & Points of Interest

The map can be found on the back cover.

1.4 Public Transportation

The closest bus and tram stops to the venue of the workshop are **Schloss** (trams: S2, S3, S9) and **Willy-Brandt-Platz** (trams: S4, S5, S6, S7, S8). Both stops are within 10 minutes walking distance to the lecture hall.

1.5 Food & Beverage

There are lots of good restaurants and bistros near TU Darmstadt. Please dial +49 6151 preceding the number given below.

Name	**Address**	**Phone**	**Cuisine**	**Opening**
Ratskeller	Marktplatz 8	26444	German	10:00 - 01:00
Pizzeria da Nino	Alexanderstr. 29	24220	Italian	18:00 - 23:00
Haroun's	Friedensplatz 6	23487	Oriental	11:00 - 22:30
Wellnitz	Lauteschlägerstr. 4	6699255	Bistro	12:00 - 24:00
Cafe Extrablatt	Marktplatz 11	5998820	Bistro	08:30 - 23:30
Ristorante Sardegna	Kahlertstraße 1	23029	Italian	11:30 - 14:45
Schwarz-Weiß-Cafe	Robert-Schneider-Str. 23	79417	Bistro	07:00 - 19:00
Schuhknecht	Schuhknechtstr. 1	4920255	Bistro	09:30 - 20:00

1.6 Conference Dinner

On Tuesday, March 17, 2026, there will be a conference dinner at the *Café Restaurant Mathildenhöhe*, Olbrichweg 13, 64287 Darmstadt, +49 172 3236231

1.7 Free Afternoon

On Wednesday, March 18, 2026, there will be a free afternoon.

Acknowledgements

Financial support by the Department of Mathematics at Justus-Liebig-Universität Gießen and the Department of Mathematics at Technische Universität Darmstadt is acknowledged.

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Mathematik

Time	Monday	Tuesday	Wednesday	Thursday	Friday
08:00	Registration				
09:00	Mondelli	Montanari	Mondelli	Montanari	Mondelli
10:30	*Coffee break*	*Coffee break*	*Coffee break*	*Coffee break*	*Coffee break*
11:00	Montanari	Mondelli	Montanari	Mondelli	Invited talk: Barbier
12:30	*Lunch*	*Lunch*	*Lunch*	*Lunch*	*Lunch*
14:00	Invited talk: Chizat	Invited talk: Döring	*Free Afternoon*	Short talks: Steiner Stonner Wu, T. Wu, Z.	*End of Spring School*
15:00	*Coffee break*	*Coffee break*		*Coffee break*	
15:30	Short talks: Maass Martínez George Schmutz Stos Fröhlich	Short talks: Schmidt Wille Chen Vasileidas Meng		Short talks: Chan Garrett Lütke Schwienhorst Tyrpak	
16:30	*Coffee break*	*Coffee break*			
17:00	Short talks: Kouraich Benning Durst Olischläger Sabelli	Short talks: Resin Schwank Tallis Kim			
	18:00 Reception	18:30 Dinner			

Monday, 16 March 2026

Time	Speaker	*Title of Talk*
09:00-09:05	Welcome	
09:05-10:35	Marco Mondelli	*Mini course* *Trustworthy machine learning through the lens of high-dim. probability*
10:35-11:00		–Coffee break–
11:00-12:30	Andrea Montanari	*Mini course* *Dynamical regimes in gradient based learning*
12:30-14:00		–Lunch break–
14:00-15:00	Lénaïc Chizat	*Invited talk* *Training dynamics of ResNets in the joint infinite width, depth and dimension limit*
15:00-15:30		–Coffee break–
15:30-15:40	Javier Maass Martínez	Short talk
15:40-15:50	Joe Bacchus George	Short talk
15:50-16:00	Valentin Schmutz	Short talk
16:00-16:10	Paul Stos	Short talk
16:10-16:20	Benedikt Fröhlich	Short talk
16:30-17:00		–Coffee break–
17:00-17:10	Anouar Kouraich	Short talk
17:10-17:20	Felix Benning	Short talk
17:20-17:30	Alison Durst	Short talk
17:30-17:40	Hans Olischläger	Short talk
17:40-17:50	Fabrizzio Sabelli	Short talk
18:00-23:00	Reception	–Cheese & Wine–

Tuesday, 17 March 2026

Time	Speaker	*Title of Talk*
09:00-10:30	Andrea Montanari	*Mini course* *Dynamical regimes in gradient based learning*
10:30-11:00		–Coffee break–
11:00-12:30	Marco Mondelli	*Mini course* *Trustworthy machine learning through the lens of high-dimensional probability*
12:30-14:00		–Lunch break–
14:00-15:00	Leif Döring	*Invited talk* *Reinforcment learning: A wonderful playing ground for probabilists*
15:00-15:30		–Coffee break–
15:30-15:40	Daniel Schmidt	Short talk
15:40-15:50	Benedikt Wille	Short talk
15:50-16:00	Ziyue Chen	Short talk
16:00-16:10	Athanasios Vasileidas	Short talk
16:10-16:20	Shanshan Meng	Short talk
16:30-17:00		–Coffee break–
17:00-17:10	Johannes Resin	Short talk
17:10-17:20	Richard Schwank	Short talk
17:20-17:30	Alexander Tallis	Short talk
17:30-17:40	Daecheol Kim	Short talk
18:30-22:00	Conference Dinner	– Café Restaurant Mathildenhöhe –

Wednesday, 18 March 2026

Time	Speaker	*Title of Talk*
09:00-10:30	Marco Mondelli	*Mini course* *Trustworthy machine learning through the lens of high-dimensional probability*
10:30-11:00		–Coffee break–
11:00-12:30	Andrea Montanari	*Mini course* *Dynamical regimes in gradient based learning*
12:30-14:00		–Lunch break–
		–Free afternoon –

Thursday, 19 March 2026

Time	Speaker	*Title of Talk*
09:00-10:30	Andrea Montanari	*Mini course* *Dynamical regimes in gradient based learning*
10:30-11:00		–Coffee break–
11:00-12:30	Marco Mondelli	*Mini course* *Trustworthy machine learning through the lens of high-dimensional probability*
12:30-14:00		–Lunch break–
14:00-14:10	Rebecca Steiner	Short talk
14:10-14:20	Jakob Stonner	Short talk
14:20-14:30	Tianqi Wu	Short talk
14:30-14:40	Zelin Wu	Short talk
15:00-15:30		–Coffee break–
15:30-15:40	Yoon Jun Chan	Short talk
15:40-15:50	Ruairi Garrett	Short talk
15:50-16:00	Lütke Schwienhorst	Short talk
16:00-16:10	Leo Tyrpak	Short talk

Friday, 20 March 2026

Time	Speaker	*Title of Talk*
09:00-10:30	Marco Mondelli	*Mini course* *Trustworthy machine learning through the lens of high-dimensional probability*
10:30-11:00		–Coffee break–
11:00-12:00	Jean Barbier	*Invited talk* *Statistical physics of deep learning near interpolation*
12:00-14:00		–Lunch, end of the Spring School–

2 List of Talks

2.1 Mini Courses

Marco Mondelli
Institute of Science and Technology (ISTA), Austria
Trustworthy machine learning through the lens of high-dimensional

This mini-course discusses theoretical guarantees in terms of robustness and privacy for over-parameterized neural networks trained via gradient descent methods, using tools from high-dimensional probability. A central focus is on the role of over-parameterization: we characterize the number of parameters p needed by neural networks to (i) interpolate n training labels, (ii) interpolate them robustly, and (iii) reconstruct n training data points (each in dimension d) from the p-dimensional trained model. As for (i), we consider linearized networks whose dynamics is accurately described by the Neural Tangent Kernel (NTK). After showing how the spectrum of the NTK (and, specifically, its smallest eigenvalue) is related to optimization [1], we give high-probability bounds on such spectrum for networks with minimal over-parameterization [2, 3], i.e., when p is roughly of the same order as n. As for (ii), we show that roughly nd parameters are both necessary and sufficient for robust interpolation [4, 5]. We then provide evidence that this same number of parameters suffices to reconstruct the whole training dataset from the trained model, addressing (iii). This inherent lack of privacy of machine learning models is well documented in the literature, and differential privacy has emerged as one of the leading paradigms to provide provable protection to any individual user. After introducing the framework of differential privacy and differentially-private versions of gradient training, we consider the random features model and show that privacy is not at odds with over-parameterization. In fact, having access to sufficiently many training data samples, differential privacy can even come with a negligible performance loss [6]. We conclude by studying the benefits of clipping gradients in the context of differentially-private linear regression, proving optimal rates by tuning the learning rate and the noise schedule [7].

References:

[1] S. Oymak and M. Soltanolkotabi. Overparameterized nonlinear learning: Gradient descent takes the shortest path?, ICML, 2019.

[2] Q. Nguyen, M. Mondelli and G. Montufar. Tight bounds on the smallest eigenvalue of the Neural Tangent Kernel for deep ReLU networks, ICML, 2021.

[3] S. Bombari, M. H. Amani and M. Mondelli. Memorization and optimization in deep neural networks with minimum over-parameterization, NeurIPS, 2022.

[4] S. Bubeck and M. Sellke. A universal law of robustness via isoperimetry, NeurIPS, 2021.

[5] S. Bombari, S. Kiyani and M. Mondelli. Beyond the universal law of robustness: Sharper laws for random features and neural tangent kernels, ICML, 2023.

[6] S. Bombari and M. Mondelli. Privacy for free in the over-parameterized regime, PNAS, 2025.

[7] S. Bombari, I. Seroussi and M. Mondelli. Better rates for private linear regression in the proportional regime via aggressive clipping, arXiv preprint, 2025.

Andrea Montanari
Stanford University, USA
Dynamical regimes in gradient based learning

Classical statistical learning theory decouples three aspects of the learning problem:

(1) Approximation (how well the model class approximates the data distribution);

(2) Generalization (the statistical error that results from learning from a finite data sample);

(3) Optimization (the algorithmic problem of minimizing the error on the training data).

Modern deep learning models are often overparametrized and trained by minimizing a non-convex empirical risk. As a consequence, the model learnt from data is not uniquely determined by the data and risk function. It is instead entirely determined by the learning dynamics and hence sensitive to its initialization, stepsize and other parameters. In these lectures I will provide a general introduction to these phenomena, and explain how different dynamical regime impact the properties of the learnt model in the case of 2-layer neural nets.

Jean Barbier
Statistical physics of deep learning near interpolation
International Center for Theoretical Physics (ICTP), Italy

For three decades, statistical physics has offered a framework to analyse neural networks, yet its applicability to deep learning with genuine feature learning remained open. I will show it is powerful enough, at least for supervised learning with a multi-layer perceptron where width scales with input dimension — a regime more expressive than narrow networks, yet more prone to feature learning than ultra-wide ones. I will focus on the interpolation regime, where parameters and data are comparable, forcing the model to adapt to the task. Working in the matched teacher-student setting will allow to identify the fundamental limits of learning and the sufficient statistics describing what an optimally trained network extracts as data increases. A rich phenomenology of learning transitions emerges. Despite its simplicity, the Bayesian-optimal analysis reveals how depth, non-linearity, and proportional width jointly shape feature learning in ways likely relevant far beyond this specific setting.

Lénaïc Chizat
Training Dynamics of ResNets in the Joint Infinite Width, Depth and Dimension Limit
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

Recent progress in AI is driven by the training of artificial neural networks of ever-increasing size. Therefore, the analysis of the infinite-size limits of such dynamics is at the core of the theoretical foundations of this discipline. While the large-width limit is by now relatively well-understood, the extension of the theory to more realistic joint limits has remained elusive. In this talk, I will present the first analysis that rigorously captures the joint infinite width, depth and (embedding) dimension limit for residual-based architectures used in practice such as the Transformer. Our approach combines techniques from stochastic approximation, propagation of chaos and dynamical mean-field theory and it leads to many insights on the structure of the randomness, on the stability of the system and on the role of the architecture's shape and hyper-parameters. We also obtain quantitative convergence rates towards the limit which are numerically tight.

Leif Döring

Reinforcment learning: A wonderful playing ground for probabilists

Universität Mannheim, Germany

One of the core topics of AI is automated decision making. While decision making is an old and classical topic in Mathematics under the name (stochastic) control theory, the point of view taken in AI is slightly different than what we are used to in Mathematics. In this talk I will work out motivations, ideas and a few algorithms that are standard in Computer Science and might perhaps be called stochastic stochastic control since we work with sampled random variables instead of solving (very non-trivial) equations. I will show how practical algorithms such as PPO and Q-learning can be framed rigorously, analysed, and improved using thoughts from probability theory.

Felix Benning
Non-stationary Isotropic Covariance Estimation
University of Luxembourg, Luxembourg

In machine learning, covariance kernels such as the Neural Tangent Kernel are typically isotropic (rotation invariant) but not stationary (translation invariant). This turns the classic theory about random fields on its head where stationary is typically assumed before isotropy is even considered. Consequently, there has been little work on non-stationary isotropic covariance kernels. In this talk we will discuss recent work on this topic: From a characterization of these kernels to an approach to estimate them from a single realization.

Yoon Jun Chan
Dynamics and spatial correlation decay for Gibbs point processes
Universität Augsburg, Germany

In a finite-size lattice spin system with continuous spin, the Gibbs measure can be seen as the invariant measure of a Glauber-Langevin dynamics with an appropriate generator. The spectral properties of the generator can be used to prove spatial correlation decay. We try to generalise this correspondence to the Gibbs point process, where there are two choices of dynamics, corresponding to diffusion and jump processes. We discuss and compare the two dynamics.

Ziyue Chen
Convergence of Entropy Regularized Log Linear Softmax Policy Gradient in Continuous Spaces: From the Perspective of Gradient Dominance
University of Edinburgh, United Kingdom

We study log-linear softmax policy gradient methods with true gradients in entropy regularized single-armed bandit problems as well as Markov Decision Process problems (MDP), which plays a fundamental role in reinforcement learning. For directly parametrized softmax policy, it is a non-convex optimization problem in the parameters, but regularization makes the algorithms converge at exponential rate while introduces bias compared to the non-regularized problem. Mei et al. [MXSS20] construct an exponential convergence rate for entropy regularized softmax policy gradient with direct parameterization in the tabular case from the perspective of the non-uniform Lojasiewicz inequality, which is a much weaker condition than convexity and inspires us to extend their ideas into more general cases. We aim to construct an exponential convergence rate for the continuous log-linear

softmax policy gradient flow using the non-uniform Polyak - Lojasiewicz inequality, in both the bandit case and MDP. Keywords: Entropy regularization, Non-convex optimization, Gradient flow, Polyak - Lojasiewicz inequality, Bandit, Markov Decision Process.
References [MXSS20] Jincheng Mei, Chenjun Xiao, Csaba Szepesvari, and Dale Schuurmans. On the global convergence rates of softmax policy gradient methods. In International conference on machine learning, pages 6820–6829. PMLR, 2020

Alison Durst
A New Correlation-Induced Bias of Mean-Decrease-in-Impurity in Random Forests in Low Signal-to-Noise-Ratio Regimes
Universität Regensburg, Germany

The mean decrease in impurity (MDI) is the canonical built-in feature importance for random forests, and its biases are generally well documented. In particular, it is known that under feature correlation, MDI can assign inflated importance to noise features that are correlated with a signal feature.
We present a different bias mechanism with the opposite qualitative implication: in regimes with low signal-to-noise-ratio and strong correlation blocks, MDI favours uncorrelated noise features over signal features and over noise features, that are correlated with the signal. We establish this effect rigorously in a synthetic setting by analysing the impurity decrease and viewing split selection as an extreme-value competition over the impurity decreases of candidate features. The key mechanism is that the block of independent noise features yield larger stochastic maxima of the impurity decrease than a highly correlated block of features, whose candidates contribute fewer effective draws. Finally, we show that local sample weighting reduces this sort of MDI bias by both theoretical results and simulation studies.

Benedikt Fröhlich
Local sample weighting increases interpretability of tree-based models for correlated features
Universität Regensburg, Germany

Feature importance statistics are a prominent and valuable tool for gaining insight into the decision process of machine learning (ML) models, but their effectiveness has well-known limitations when features within the training data are correlated. We propose local sample weighting, which can be integrated locally within the split point selection process of tree based estimators to obtain better estimates of the marginal effects of each feature. Our approach uses estimated density ratio weights to decorrelate a target feature from the remaining features. This reduces model bias locally, whenever the effect of a feature is

evaluated and compared to others. Moreover, it comes with a natural tuning parameter, the minimum effective sample size of the weighted population, which corresponds to an interpretation-prediction-tradeoff, analog to a biasvariance-tradeoff as for classical ML tuning parameters.
We prove consistency for our modified tree estimator and show that it comes with a notion of mean decrease of impurity (MDI) importance that asymptotically distinguishes between signal and noise features.

Ruairi Garrett
Genealogy of large logistic branching processes
University of Oxford, United Kingdom

Logistic branching processes have been introduced as a simplified model of a population competing for resources. Genetic drift in such populations was recently studied (Forien '25) and shown to converge to the family of so called Beta-Fleming-Viot processes. In this talk we will give some background on (probabilistic) population genetics, and discuss an extension of the result of Forien to obtain the behaviour of the genealogy relating finite samples taken from the population at some fixed time horizon. These are shown to converge to Kingman, Beta (2-alpha, alpha) and Bolthausen-Sznitman coalescents. Based on joint work with Julio Ernesto Nava Trejo.

Joe Bacchus George
Neural Belief Propagation for Combinatorial Optimisation
University of Cambridge, United Kingdom

Combinatorial optimisation is fundamentally challenging, and progress relies on our ability to find effective heuristics. Graph neural networks (GNNs) are one proposed heuristic. These methods work by iteratively passing "messages" between the nodes of a graph and using a neural network to parameterize the update rules. To date, GNNs have mostly failed to achieve competitive performance. Belief propagation is an alternative heuristic that is particularly effective on locally tree-like graphs. The method also works by iteratively passing messages between the nodes of a graph but, in contrast, the update equations are derived a priori and have no free parameters.
We consider a hybrid approach that supplements the belief propagation equations with a GNN. The neural network is trained unsupervised on small graphs to find corrections to the original belief propagation equations. We present systematic experimental results for the hybrid approach across a range of problems. We find that the combined method considerably outperforms both constituent parts. At the same time, the combined method still underperforms simulated annealing in difficult instances. Our results highlight that

building strong heuristics directly into GNN architectures vastly improves performance. The corollary is that GNNs are fundamentally limited by our knowledge of conventional non-learned heuristics and thus our ability to design effective architectures.

Daecheol Kim
Phase transitions in scale-free long-range first passage percolation on lattices
University of Illinois Urbana–Champaign, USA

First passage percolation (FPP) is a classical model for random growth and propagation on networks. In this talk, I introduce a scale-free long-range FPP model on d-dimensional lattices, where passage times depend on both geometric distance and heavy-tailed randomness. The model extends the long-range FPP framework of Chatterjee and Dey by allowing both general edge-weight distributions beyond the exponential case and a scale-free structure through vertex weights, leading to a rich phase diagram for first passage times. Conceptually, it combines a spatial random metric induced by the long-range lattice geometry with a geometry-independent scale-free inhomogeneity arising from heavy-tailed vertex weights. I will focus on the resulting phase transitions in the leading-order behavior.
This is joint work with Shirshendu Chatterjee and Partha S. Dey

Anouar Kouraich
An introduction to the limit behaviour of Quantum P-Spin Glasses for large p
TU München, Germany

I present some results on the spectral properties of Quantum Ising p-spin glasses. At high p, this model converges to the Quantum Random Energy Model. I will then shortly discuss the tools and difficulties involved in analysing these quantum spin glasses compared to their classical counterparts. This is accomplished by combining existing analytical techniques that address the non-commutative properties of such quantum glasses with a description of the typical geometry of extreme negative deviations in the classical spin glass. If I have enough time, I will also review properties conjectures addressing corrections in the quantum case.

Claudius Lütke Schwienhorst
Lévy Langevin Monte Carlo
TU Dresden, Germany

We extend the Monte Carlo method of (Oechsler 2024) in several ways. Replacing the classical Langevin diffusion with the solution of a generalised Lévy-driven Langevin equation

allows us to overcome its shortcomings when sampling from a given distribution via its unnormalised density. We show exponential convergence in case of multimodal, heavy-tailed and non-smooth probability distributions on $\mathbb{R}^d$, which in general cannot be guaranteed by the classical Langevin diffusion. In particular, we implement the method with a compound Poisson process as driving noise. This choice is motivated by the fact that the convergence of the continuous time SDE then extends favorably to the actual implementation. The reason for this is twofold: For one, the noise term can be simulated exactly for finite intensity jumps, resulting in numerical benefits for a discretisation. Second, the nonlocal terms appearing in the drift function may be efficiently approximated via Monte Carlo methods due to the finite intensity of the jumps.
References:
Oechsler, D.: Lévy Langevin Monte Carlo. *Stat. Comput.*, 34 (37): 1-15, 2024.

Javier Maass Martínez
ResNets of All Shapes and Sizes: Quantitative Large-Scale Theory of Training Dynamics
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

We study the convergence of the training dynamics of residual neural networks (ResNets) towards their joint infinite depth-width-embedding limit. Specifically, we consider ResNets with two-layer perceptron blocks, whose shape is determined by their depth L, hidden width M and embedding dimension D and we adopt the residual scaling $O(\frac{\sqrt{D}}{ML})$ identified in [Chizat, 2025] as necessary for "local" feature learning. We prove that, after a bounded number of training steps, the error between the ResNet and its infinite-size limit is $O\big(\frac{1}{L} + \frac{\sqrt{D}}{\sqrt{LM}} + \frac{1}{\sqrt{D}}\big)$. Our analysis applies formally to a broad class of residual architectures, including Transformers with bounded key/query dimension. From a technical viewpoint, we take the large-D limit of the Mean ODE model introduced in [Chizat, 2025], and establish convergence at a rate $O(1/\sqrt{D})$. The limit dynamics is of Dynamical Mean Field Theory (DMFT)-type and inherits a rich probabilistic structure from the random initialization. To handle this structure, we combine the cavity method with propagation of chaos arguments at a functional level on so-called *skeleton maps* which we introduce. Note: This is ongoing work, so the title and abstract might be slightly modified in the future.

Shanshan Meng
Supervised classification for Ornstein-Uhlenbeck diffusions with separation condition
HU Berlin, Germany

We study binary supervised classification based on repeated independent observations of continuous sample paths. Our focus is a diffusion classification model in which the features

follow an Ornstein-Uhlenbeck process with class-dependent drifts. We consider plug-in classifiers constructed from drift estimators and analyze the performance via the excess risk.
Under a separation condition on the drift parameters, we establish upper bounds of the excess risk, which are explicitly parametrized by the separation distance quantifying the difficulty of the problem. Specifically, when the drift distance is bounded away from zero, the plug-in classifiers achieve a fast convergence rate of order 1/n (up to logarithmic factors) in the constant drift scenario. Furthermore, we discuss extensions of this framework to time-inhomogeneous drift functions.

Hans Olischläger
Directions of complexity - conditional distributions in Amortized Bayesian Inference
TU Dortmund, Germany

Amortized Bayesian Inference is an approach to Bayesian computation where a statistical (mostly neural) model of a conditional probability distributions is learned to solve analytically intractable scientific inverse problems from physics to psychology. To match the trustworthiness of gold-standard MCMC approaches, theoretical and empirical understanding of proper scaling of model to data complexity is lacking. I will present ongoing work on systematic study of the bias-variance trade-off for normalizing flows, flow matching and direct feed forward point estimation architectures for different inverse problems and argue for decomposing model complexity into two parts: (1) complexity of the learn distribution for a fixed condition and (2) complexity of the dependence on the condition.

Johannes Resin
Score decompositions and calibration error for probabilistic classifiers
Goethe-Universität Frankfurt, Germany

Decompositions of proper scores into measures of miscalibration (reliability), discrimination (resolution), and uncertainty have a long history in weather forecasting. In contrast, calibration error metrics have only been of relatively recent interest in machine learning. In this short talk, I highlight the close connection between calibration error metrics and the miscalibration term of score decompositions. A small case study on common image classifiers shows that a sole focus on calibration error can produce misleading conclusions, as it may substantially deteriorate predictive performance.

Fabrizzio Sabelli
High-dimensional limit of streaming SGD for ℓ_2-adversarially trained generalized linear models
Université de Montréal, Canada

Adversarial training is used to increase the robustness of machine learning models towards attacks by malicious actors. We characterize the high-dimensional limit of adversarial training with respect to the ℓ_2-norm of generalized linear models in the setting of streaming, single batch stochastic gradient descent (SGD) when the number of samples scales proportionally with the problem dimension. We characterize the high-dimensional limit of SGD as a stochastic differential equation, denoted adversarial homogenized SGD. From this limit, we derive a deterministic system of ordinary differential equations which provides insights into the generalization capabilities of SGD.

Daniel Schmidt
How Finite Geometric Sums Fix a Boundary Effect in RL
Universität Mannheim, Germany

Reinforcement learning trains agents to make sequential decisions by interacting with an environment and relies on low-variance estimators of action quality. Generalized Advantage Estimation is widely used and can be viewed as a geometrically weighted mixture of k-step bootstrapped return estimators. While in theory the geometric weights are normalized over an infinite horizon, in practice rollouts are finite and the mixture must be truncated at the end of trajectories. This induces a previously overlooked boundary effect. Missing tail mass is shifted to the longest available k-step term, changing the bias-variance trade-off near trajectory ends. A normalized finite geometric sum removes this artifact while retaining a simple backward recursion. The resulting estimator can speed up learning, as demonstrated in a simple experiment.

Valentin Schmutz
High-dimensional neuronal activity from low-dimensional latent dynamics: a solvable model
University of Oxford, United Kingdom

Computation in recurrent networks of neurons has been hypothesized to occur at the level of low-dimensional latent dynamics, both in artificial systems and in the brain. This hypothesis seems at odds with evidence from large-scale neuronal recordings in mice showing that neuronal population activity is high-dimensional. To demonstrate that low-dimensional latent dynamics and high-dimensional activity can be two sides of the same coin, we present an analytically solvable recurrent neural network (RNN) model whose dynamics can be exactly reduced to a low-dimensional dynamical system, but generates an activity manifold that has a high linear embedding dimension. Our model allows us to combine the theory of mean-field low-rank recurrent neural networks (from computational neuroscience) and the theory of random feature neural networks (from machine learning), and suggests that theoretical results on the latter might be relevant for recurrent neural networks as well.

Richard Schwank
Robust Score Matching
TU München, Germany

Score matching is a parameter estimation procedure that does not require computation of distributional normalizing constants. Relying on the geometric median of means, we construct a score matching procedure for possibly very high-dimensional exponential families that yields consistent parameter estimates even if the observed data has been partly contaminated. As a side-product, we find the bias of the geometric median for certain high-dimensional distributions to be much smaller than previously known.

Rebecca Steiner
Multi-drawing Pólya urns via labelled random DAGs
JGU Mainz, Germany

A Pólya urn of replacement matrix $R = (R_{i,j})_{1\le i,j\le d}$ is a Markov process that encodes the following experiment: an urn contains balls of d different colours and at every time-step, a ball is drawn uniformly at random in the urn, and if its colour is i, then it is replaced in the urn with an additional $R_{i,j}$ balls of colour j, for all $1 \le i, j \le d$. We study a natural extension of this model in which, instead of drawing one ball at each time-step, we draw a set of $m \ge 2$ balls: in this case, the replacement matrix becomes a replacement tensor. Because of the multi-draws, this process can no longer be seen as a branching process, which makes its analysis much more intricate than in the classical Pólya urn case. Partial results proved by stochastic approximation techniques exist in the literature. In this article, we introduce a new approach based on seeing the process as a stochastic process indexed by a random directed-acyclic graph (DAG) and use this approach, together with the theory of stochastic tensors, to prove a convergence theorem for these multi-drawing Pólya urns, with assumptions that are straightforward to check in practice.

Jakob Stonner
A principle of one big jump for the branching random walk
JLU Gießen, Germany

We prove a version of Nagaev's theorem for the branching random walk with heavy-tailed associated random walk. For a heavy-tailed random walk $(S_n)_{n\in\mathbb{N}}$ on $\mathbb{R}$, this theorem roughly states that $\mathbb{P}(S_n > n\mathbb{E}[S_1] + t) = n\mathbb{P}(S_1 > t)(1 + o(1))$ as $n \to \infty$ uniformly in large $t > 0$, which is explained probabilistically by the principle of one big jump. For a branching random walk on $\mathbb{R}$ we consider the random measure $\widehat{Z}_n = \sum_{|u|=n} e^{-V_u}\delta_{V_u}$, where $n \in \mathbb{N}$ and V_u, $|u| = n$ denote the positions of the particles in the n-th generation. Our result then states that $\widehat{Z}_n([n\mathbb{E}[S_1] + t_n, \infty)) = Wn\mathbb{P}(S_1 > t_n)(1 + o(1))$ in L^1 as $n \to \infty$, where W is a non-degenerate random variable, $t_n \uparrow \infty$ grows fast enough, and $(S_n)_{n\in\mathbb{N}}$ is the random walk with law $\mathbb{E}[\widehat{Z}_1(\cdot)]$.

Paul Stos
Fluctuations and Concentration in Two-layer Neural Networks
Université Clermont Auvergne, France

We study the learning dynamics of wide two-layer neural networks trained by stochastic gradient descent (SGD), aiming to understand quantitatively how width shapes both the typical training trajectory and the variability of the final predictor.
We adopt an interacting particle viewpoint in which N neurons evolve under SGD as a large coupled system. As the number of neurons grows, this collective dynamics is well approximated by a deterministic mean-field limit, which provides a more tractable description of how the parameter distribution (and hence predictions) evolves during training.
We then quantify finite-width effects through two complementary results. First, we characterise fluctuations around the mean-field limit: after the natural $\sqrt{N}$ rescaling, deviations converge to a Gaussian limiting process, yielding an explicit description of the variability induced by training randomness. Second, we establish finite-width concentration inequalities–uniform over training time–controlling with high probability how close a width-N network remains to its mean-field proxy

Alexander Tallis
Elucidating Noise Scheduling Methods for Gaussian Probability Paths
University of Bristol, United Kingdom

Flow and Diffusion Based Models have emerged as the dominant tool in many fields of generative AI, primarily in image generation. These models rely on the so called 'probability path' which describes the process in which data is corrupted into noise. Gaussian Probability Paths are commonly chosen in practice due to their simplicity and closed form solutions. We explore the different choices of probability path available and observe close connections between noise scheduler selection and sampling distribution in model design.

Leo Tyrpak
Fluctuations of spatial population models with non-local interactions.
University of Oxford, United Kingdom

We analyse a class of spatial population models where the branching rate of particles depends non-locally on the whole empirical measure of particles. The scaling limit of this model has previously been shown to be the non-local Fisher-KPP equation. In this talk we will discuss the fluctuations of the process around these deterministic limits and show how these can tell us about ancestry.

Athanasios Vasileidas
Markov Decision Processes of the Third Kind: Learning Distributions by Policy Gradient Descent
KIT Karlsruhe, Germany

In this talk we will introduce and analyze a Policy Gradient algorithm for distributional Markov Decision Processes as a class of control problems in which the objective is to learn policies that steer the distribution of a cumulative reward toward a prescribed target law, rather than optimizing an expected value or a risk functional. We prove convergence of the algorithm to stationary points using stochastic approximation techniques. Several numerical experiments illustrate the ability of the method to match complex target distributions, recover classical optimal policies when they exist, and reveal intrinsic non-uniqueness phenomena specific to distributional control.

Benedikt Wille
The role of Target Update Frequencies in Q-Learning
Universität Mannheim

Target networks are a celebrated milestone enabling the massive success of deep Temporal Difference learning methods such as Deep Q Networks due to their stabilizing effect. Despite their popularity, the central decision problem of choosing the target network update frequency (TUF) remains poorly understood and it is often treated as merely another tunable hyperparameter rather than as a principled design decision. We characterize periodic target updates in Q-learning as a nested optimization scheme in which each outer iteration applies an inexact Bellman optimality operator approximated by a generic inner loop optimizer. Our analysis yields optimal TUF schedules when applying stochastic gradient descent to the inner loop optimization problem.

Tianqi Wu

2nd order asymptotics for the hard wall probability of the 2D harmonic crystal

Technion Haifa, Israel

We estimate the probability that the discrete Gaussian free field on a planar domain with Dirichlet boundary conditions stays positive in the bulk. Improving upon the result by Bolthausen, Deuschel and Giacomin from 2001, we derive the order of the subleading term of this probability when a sequence of discretized scale-ups of given domain and compactly included smooth bulk are considered. A main ingredient in the proof is the double exponential decay of the right tail of the centered minimum of the field in the bulk, conditioned on a certain weighted average of its values to be zero. (Joint work with Maximilian Fels and Oren Louidor.)

Zelin Wu

PINNs for Rough Volatility Models

Universität Bonn, Germany

Rough volatility models have demonstrated clear advantages over classical stochastic volatility models in accurately capturing observed volatility surfaces. However, they are computationally challenging due to the non-Markovian nature of the fractional Brownian motion. We employed a Markovian approximation to bring rough models into a high-dimensional Markovian framework. Then we can apply Physics-Informed Neural Networks (PINNs) to find efficient numerical solutions.

3 List of Participants

Aurzada, Frank TU Darmstadt, Germany

Barbier, Jean International Centre for Theoretical Physics (ICTP), Italy

Benning, Felix University of Luxembourg, Luxembourg

Betz, Volker TU Darmstadt, Germany

Birkner, Matthias JGU Mainz, Germany

Casarin, Luca University of Padova, Italy

Chan, Yoon Jun Universität Augsburg, Germany

Chen, Ziyue University of Edinburgh, United Kingdom

Chizat, Lénaïc École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

De Marzi, Marco University of Padova, Italy

Döring, Leif Universität Mannheim, Germany

Durst, Alison Universität Regensburg, Germany

Eickmeyer, Kord TU Darmstadt, Germany

Franck, Sascha Universität Göttingen, Germany

Fröhlich, Benedikt Universität Regensburg, Germany

Garrett, Ruairi University of Oxford, United Kingdom

George, Joe Bacchus University of Cambridge, United Kingdom

Gerisch, Alf TU Darmstadt, Germany

Helmer, Max TU Darmstadt, Germany

Huang, Kai-Chun Eindhoven University of Technology, Netherlands

Iyer, Tejas WIAS Berlin, Germany

Junike, Gero LMU München, Germany

Kampf, Jürgen Universitätsklinikum Essen, Germany

Kim, Daecheol University of Illinois Urbana–Champaign, USA

Klippel, Andreas TU Darmstadt, Germany

Kohler, Michael TU Darmstadt, Germany

Kolodziejska, Alicja JLU Gießen, Germany

Köppl, Heinz TU Darmstadt, Germany

Kouraich, Anouar TU München, Germany

Kraft, Mino Nicola TU Darmstadt, Germany

Kuhl, Petra JLU Gießen, Germany

Liu, Jiachi Goethe-Universität Frankfurt, Germany

Lütke Schwienhorst, Claudius TU Dresden, Germany

Maass Martínez, Javier École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

Meiners, Matthias JLU Gießen, Germany

Meng, Shanshan HU Berlin, Germany

Molinero Römer, Vincent TU Darmstadt, Germany

Mondelli, Marco Institute of Science and Technology Austria (ISTA), Austria

Montanari, Andrea Stanford University, USA

Müller, Sabine TU Darmstadt, Germany

Olischläger, Hans TU Dortmund, Germany

Resin, Johannes Goethe-Universität Frankfurt, Germany

Sabelli, Fabrizzio Université de Montréal, Canada

Schmidt, Daniel Universität Mannheim. Germany

Schmidt, Marwin University College London, United Kingdom

Schmidt, Tobias TU Darmstadt, Germany

Schmutz, Valentin University of Oxford, United Kingdom

Schwank, Richard TU München, Germany

Steiner, Rebecca JGU Mainz, Germany

Stonner, Jakob JLU Gießen, Germany

Stos, Paul Université Clermont Auvergne, France

Tallis, Alexander University of Bristol, United Kingdom

Tarasov, Aleksandr Universität Hildesheim, Germany

Teichmann, Kaya Emelie Universität Hildesheim, Germany

Thaller, Tobias Universität Heidelberg, Germany

Torrisi, Giovanni Luca CNR Rome, Italy

Tuchel, Philipp Ruhr-Universität Bochum, Germany

Tyrpak, Leo University of Oxford, United Kingdom

Vasileidas, Athanasios KIT Karlsruhe, Germany

Verovkin, Glib Universität Hildesheim, Germany

Wiegand, Max TU Darmstadt, Germany

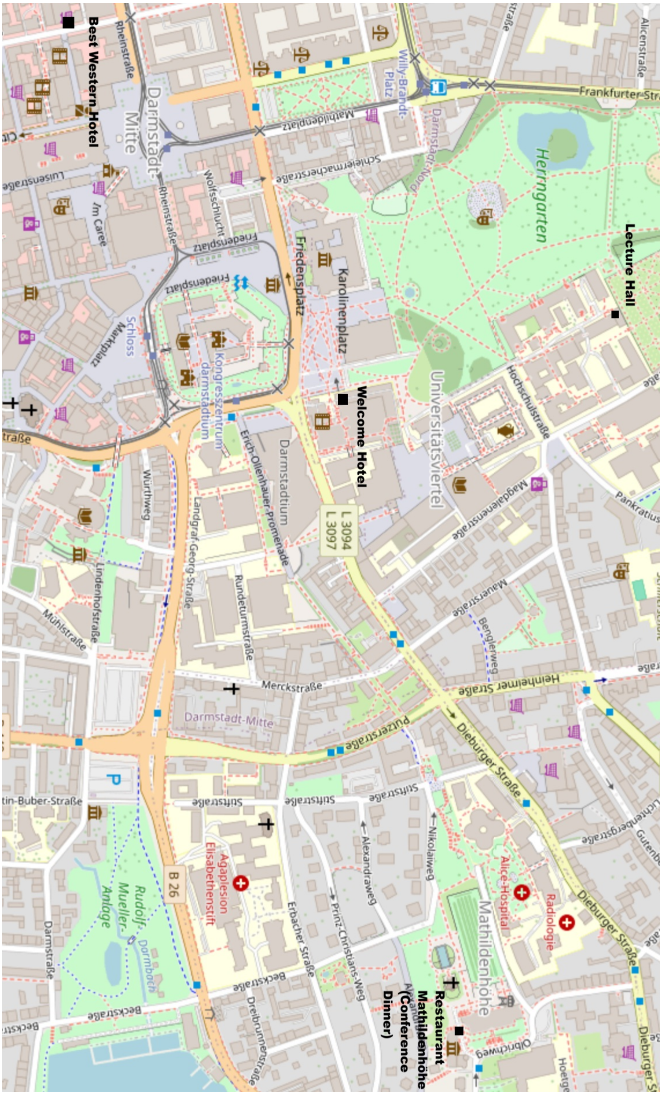
Wille, Benedikt Universität Mannheim, Germany

Worf, Virginia TU Darmstadt, Germany

Wu, Tianqi Technion Haifa, Israel

Wu, Zelin Universität Bonn, Germany

Zarifian, Arefeh TH Mittelhessen /JLU Gießen, Germany


Best Western Hotel
Welcome Hotel
Lecture Hall
Restaurant Mathildenhöhe (Conference Dinner)
Darmstadt-Mitte
Herrngarten
Universitätsviertel
Darmstadium
Kongresszentrum darmstadium
Schloss
Friedensplatz
Karolinenplatz
Mathildenplatz
Luisenstraße
Rheinstraße
Frankfurter Straße
Dieburger Straße
Heinheimer Straße
Pützerstraße
Stiftstraße
Merckstraße
Landgraf-Georg-Straße
Erich-Ollenhauer-Promenade
Magdalenenstraße
Hochschulstraße
Mathildenhöhe
Alice-Hospital
Agaplesion Elisabethenstift
Rudolf-Mueller-Anlage
Beckstraße
Erbacher Straße
B 26
L 3094
L 3097