丘成桐数学科学中心
YAU MATHEMATICAL SCIENCES CENTER

TSIMF
清华三亚国际数学论坛

统计人工智能
Statistical AI



会议时间 TIME
August 24-28, 2026

会议地点 VENUE
Room A-103, TSIMF

组织者 ORGANIZERS

Eric Moulines, MBZUAI, Ecole Polytechnique
Alexey Naumov, HSE University, Steklov Mathematical Institute of RAS
Sergey Samsonov, HSE University
Yuhao Wang (王禹皓), Tsinghua University

Contents

About the conference 1

Schedule 5

August 24, 2026 - Monday 7

August 25, 2026 - Tuesday 8

August 26, 2026 - Wednesday 9

August 27, 2026 - Thursday 10

August 28, 2026 - Friday 11

Titles and Abstracts 12

Michal Valko 13

Vladimir Spokoiny 13

Bingyi Jing(荆炳义) 14

Kenji Fukumizu 14

Nikita Puchkin 15

Praneeth Vepakomma 16

Alex Lamb(杨丁怀) 16

Tengyao Wang(王藤耀) 17

Yunbei Xu(许云贝) 17

Konstantin Yakovlev 18

Gen Li(李根) 18

Miha Bresar 19

Hoi-To Wai(韋凱滔) 20

Denis Belomestny 20

Zhiheng Zhang(张智恒) 21

Junpei Komiyama 21

Lei Wu(吴磊) 22

Marina Sheshukova 22

Maxim Rakhuba 23

Welcome to TSIMF 24

About Facilities 26

Registration 26

Guest Room 26
Library 27
Restaurant 27
Laundry 28
Gym 28
Swimming Pool 28
Free Shuttle Bus Service at TSIMF 28
Playground 28
Contact Information of Administration Staff 29

About the conference



统计人工智能研讨会
Statistical AI workshop

Date
August 24-28, 2026

Venue
Room A-103, TSIMF

Organizers
Eric Moulines, MBZUAI, Ecole Polytechnique

Alexey Naumov, HSE University, Steklov Mathematical Institute of RAS

Sergey Samsonov, HSE University

Yuhao Wang (王禹皓), Tsinghua University

Abstract
The Statistical AI workshop focuses on the statistical ideas that make modern AI reliable - how to quantify uncertainty, make decisions under uncertainty, and design theoretically grounded algorithms that still scale to large data settings. As models grow larger and are deployed more widely, these questions become both more urgent and more technically challenging. The workshop will bring together researchers across statistics, machine learning, and applied AI to discuss recent results and open problems. The workshop will cover, but is not limited to, the following key topics:

- Statistical decision making
- Generative modelling and sampling
- Conformal prediction
- Statistical and causal inference
- Stochastic optimisation, stochastic approximation, and optimization methods for large-scale and non-convex AI models

Participants

1. Miha Bresar, The Chinese University of Hong Kong, China
2. Ziyi Chen(陈子怡), Southern University of Science and Technology, China
3. Kenji Fukumizu, The Institute of Statistical Mathematics, Japan
4. Wenjie Huang(黄文杰), Tsinghua University, China
5. Binyi Jing(荆炳义), The Chinese University of Hong Kong, Shenzhen, China
6. Junpei Komiyama, Mohammed Bin Zayed University of Artificial Intelligence, UAE
7. Chenyu Kong(孔晨宇), Tsinghua University, China
8. Alex Lamb(杨丁怀), Tsinghua University, China
9. Ilya Levin, HSE University, Russia
10. Gen Li(李根), The Chinese University of Hong Kong, China
11. Shaojie Li(李少杰), National University of Singapore, Singapore
12. Zhiyi Li(李智翼), National University of Singapore, Singapore
13. Zonghan Li(李宗翰), Tsinghua University, China
14. Yuhao Liu(刘雨濠), Tsinghua University, China
15. Ian Maximov, HSE University, Russia
16. Nikita Morozov, HSE University, Russia
17. Eric Moulines, MBZUAI, Ecole Polytechnique, UAE / France
18. Alexey Naumov, HSE University, Steklov Mathematical Institute of RAS, Russia
19. Chung Nguyen, National University of Singapore, Singapore
20. Nikita Puchkin, HSE University, Russia
21. Maxim Rakhuba, HSE University, Russia
22. Sergey Samsonov, HSE University, Russia
23. Marina Sheshukova, HSE University, Russia
24. Jiaxin Shi(师佳鑫), Peking University, China
25. Vladimir Spokoiny, HSE University, Russia; WIAS Berlin, Germany
26. Denis Suchkov, HSE University, Russia
27. Vladimir Ulyanov, HSE University, Russia
28. Michal Valko, Institut national de recherche en informatique et en automatique(INRIA), France
29. Praneeth Vepakomma, MIT, USA; MBZUAI, UAE
30. Hoi-To Wai(韋凱滔), The Chinese University of Hong Kong, China
31. Tengyao Wang(王藤耀), London School of Economics, UK
32. Wei Wang(王炜), Nanjing University of Aeronautics and Astronautics, China

33. Yuhao Wang(王禹皓), Tsinghua University, China
34. Lei Wu(吴磊), Peking University, China
35. Yunbei Xu(许云贝), National University of Singapore, Singapore
36. Konstantin Yakovlev, HSE University, Russia
37. Wei Yao(姚巍), Renmin University of China, China
38. Yuzhe Yuan(元郁哲), National University of Singapore, Singapore
39. Jingru Zhang(张静茹), Fudan University, China
40. Zhiheng Zhang(张智恒), Shanghai University of Finance and Economics, China



Schedule

Time&Date	Monday (August 24)	Tuesday (August 25)	Wednesday (August 26)	Thursday (August 27)	Friday (August 28)
7:30-8:30	*Breakfast (60 minutes)*				
Chair	Alexey Naumov	Alexey Naumov	2 talks by PhD students:	Nikita Puchkin	Nikita Puchkin
9:00-10:00	Michal Valko	Michal Valko	9:00-9:25 Jiaxin Shi 9:30-10:00 Wei Yao	Gen Li	Gen Li
10:00-10:30	*Coffee Break (within 30 minutes)*				
Chair	Yuhao Wang	Gen Li	3 talks by PhD students:	Sergey Samsonov	Sergey Samsonov
10:30-11:10	Vladimir Spokoiny	Kenji Fukumizu	10:30-10:55 Yuzhe Yuan 11:00-11:25 Nikita Morozov 11:30-11:55 Ilya Levin	Miha Bresar	Maxim Rakhuba
11:15-11:55	Bingyi Jing	Nikita Puchkin		Hoi-To Wai	
12:00-13:30	*Lunch (90 minutes)*				
Chair	*Free Discussion*	Gen Li	*Free Discussion*	Jing Binyi	*Free Discussion*
13:30-14:10		Praneeth Vepakomma		Denis Belomestny	
14:15-14:55		Alex Lamb		Zhiheng Zhang	
15:00-15:30		*Coffee Break*		*Coffee Break*	
Chair		Alexey Naumov		Jing Binyi	
16:00-16:10		Tengyao Wang		Junpei Komiyama	
16:15-16:55		Yunbei Xu		Lei Wu	
17:00-17:40		Konstantin Yakovlev		Marina Sheshukova	
		Group photo (17:40-18:00)			
18:00	*Dinner (90 minutes)*	*Banquet 18:00-20:00*	*Dinner (90 minutes)*		

August 24, 2026 - Monday

Time	Name	Title
7:30-8:30	*Breakfast (60 minutes)*	
Chair	**Alexey Naumov** HSE University, Steklov Mathematical Institute of RAS	
9:00-10:00	**Michal Valko** Institut national de recherche en informatique et en automatique	The genesis of predictive architectures
10:00-10:30	*Coffee Break (within 30 minutes)*	
Chair	**Yuhao Wang(王禹皓)** Tsinghua University	
10:30-11:10	**Vladimir Spokoiny** HSE University, WIAS Berlin	Manifold Gaussian Mixture Models
11:15-11:55	**Bingyi Jing(荆炳义)** The Chinese University of Hong Kong, Shenzhen	Balancing Intelligence and Efficiency in Large Language Models
12:00-13:30	*Lunch (90 minutes)*	
Chair	*Free Discussion*	
13:30-14:10		
14:15-14:55		
15:00-15:30		
Chair		
15:30-16:10		
16:15-16:55		
17:00-17:40		
18:00	*Dinner (90 minutes)*	

August 25, 2026 - Tuesday

Time	Name	Title
7:30-8:30	*Breakfast (60 minutes)*	
Chair	**Alexey Naumov** HSE University, Steklov Mathematical Institute of RAS	
9:00-10:00	**Michal Valko** Institut national de recherche en informatique et en automatique	The genesis of predictive architectures
10:00-10:30	*Coffee Break (within 30 minutes)*	
Chair	**Gen Li(李根)** The Chinese University of Hong Kong	
10:30-11:10	**Kenji Fukumizu** The Institute of Statistical Mathematics	Flow Matching from Viewpoint of Proximal Operator
11:15-11:55	**Nikita Puchkin** HSE University	Stochastic optimal control approach to generative modelling and Schrödinger potential estimation
12:00-13:30	*Lunch (90 minutes)*	
Chair	**Gen Li(李根)** The Chinese University of Hong Kong	
13:30-14:10	**Praneeth Vepakomma** MIT, MBZUAI	Modulated learning for private and distributed regression with just a single sample per client device
14:15-14:55	**Alex Lamb(杨丁怀)** Tsinghua University	Next-Latent Prediction Transformers Learn Compact World Models
15:00-15:30	*Coffee Break (within 30 minutes)*	
Chair	**Alexey Naumov** HSE University, Steklov Mathematical Institute of RAS	
15:30-16:10	**Tengyao Wang(王藤耀)** London School of Economics	Optimal in-context adaptivity and distributional robustness of transformers
16:15-16:55	**Yunbei Xu(许云贝)** National University of Singapore	Pointwise Generalization in Deep Neural Networks
17:00-17:40	**Konstantin Yakovlev** HSE University	Generalization bounds for Diffusion Models and Beyond
17:40-18:00	*Group photo*	
18:00	*Banquet (18:00-20:00)*	

August 26, 2026 - Wednesday

Time	Name	Title
7:30-8:30	*Breakfast (60 minutes)*	
9:00-10:00	**2 talks by PhD students:** 9:00-9:25 - **Jiaxin Shi(师佳鑫)** 9:30-10:00 - **Wei Yao(姚巍)**	
10:00-10:30	*Coffee Break (within 30 minutes)*	
10:30-11:55	**3 talks by PhD students:** 10:30-10:55 - **Yuzhe Yuan(元郁哲)** 11:00-11:25 - **Nikita Morozov** 11:30-11:55 - **Ilya Levin**	
12:00-13:30	*Lunch (90 minutes)*	
13:30-14:10	*Free Discussion*	
14:15-14:55		
15:00-15:30		
15:30-16:10		
16:15-16:55		
17:00-17:40		
18:00	*Dinner (90 minutes)*	

August 27, 2026 - Thursday

Time	Name	Title
7:30-8:30	*Breakfast (60 minutes)*	
Chair	**Nikita Puchkin** HSE University	
9:00-10:00	**Gen Li(李根)** The Chinese University of Hong Kong	Foundations of Diffusion Generative Models
10:00-10:30	*Coffee Break (within 30 minutes)*	
Chair	**Sergey Samsonov** HSE University	
10:30-11:10	**Miha Bresar** The Chinese University of Hong Kong, Shenzhen	Diffeomorphic Markov Chain Monte Carlo: fast mixing for heavy-tailed distributions
11:15-11:55	**Hoi-To Wai(韋凱滔)** The Chinese University of Hong Kong	Stochastic Approximation Schemes with Decision-Dependent Samples: The Case of Performative Prediction
12:00-13:30	*Lunch (90 minutes)*	
Chair	**Bingyi Jing(荆炳义)** The Chinese University of Hong Kong, Shenzhen	
13:30-14:10	**Denis Belomestny** HSE University, Duisburg Essen University	Statistical analysis of Inverse Entropy-regularized Reinforcement Learning
14:15-14:55	**Zhiheng Zhang(张智恒)** Shanghai University of Finance and Economics	Wasserstein Policy Learning for Distributional Outcomes
15:00-15:30	*Coffee Break (within 30 minutes)*	
Chair	**Bingyi Jing(荆炳义)** The Chinese University of Hong Kong, Shenzhen	
15:30-16:10	**Junpei Komiyama** Mohammed Bin Zayed University of Artificial Intelligence	Statistical Certification for Majority Voting
16:15-16:55	**Lei Wu(吴磊)** Peking University	Functional Scaling Laws for LLM Pretraining
17:00-17:40	**Marina Sheshukova** HSE University	Gaussian Approximation and Bootstrap Inference for Stochastic Gradient Methods and Reinforcement Learning
18:00	*Dinner (90 minutes)*	

August 28, 2026 - Friday

Time	Name	Title
7:30-8:30	*Breakfast (60 minutes)*	
Chair	**Nikita Puchkin** HSE University	
9:00-10:00	**Gen Li(李根)** The Chinese University of Hong Kong	Foundations of Diffusion Generative Models
10:00-10:30	*Coffee Break (within 30 minutes)*	
Chair	**Sergey Samsonov** HSE University	
10:30-11:10	**Maxim Rakhuba** HSE University	Linear Algebra Problems and Algorithms in Muon-based Optimization Methods
11:15-11:55		
12:00-13:30	*Lunch (90 minutes)*	
Chair	*Free Discussion*	
13:30-14:10		
14:15-14:55		
15:00-15:30		
Chair		
15:30-16:10		
16:15-16:55		
17:00-17:40		
18:00	*Dinner (90 minutes)*	

Titles and Abstracts



The genesis of predictive architectures

Michal Valko
Institut national de recherche en informatique et en automatique(INRIA)

Lecture 1: In this talk, we will discuss self-supervised representation learning and then more specifically BYOL. BYOL relies on two neural networks, referred to as online and target networks, that interact and learn from each other. From an augmented view of an image, we train the online network to predict the target network representation of the same image under a different augmented view. At the same time, we update the target network with a slow-moving average of the online network. While state-of-the art methods had intrinsically relied on negative pairs, BYOL achieved a new state of the art without them. We will also describe follow-ups of BYOL that we have explored within DeepMind, BGRL (for graphs), MYOW (for new uncharted domains such as neural readings), and BAS/BEAST (for multi-modal domains). Finally, we will explore different takes on the analysis of BYOL and other open questions.

Lecture 2: Reinforcement Learning from Human Feedback (RLHF) typically fits a reward model (often Bradley-Terry) and then optimizes a policy (e.g., PPO or DPO). These assumptions can be distribution-dependent and too restrictive to reflect the full structure of human preferences, including non-transitive behavior. We present Nash Learning from Human Feedback (NLHF), which uses pairwise comparisons to induce a two-player preference game and frames alignment as computing a Nash equilibrium, avoiding an explicit reward model. The talk focuses on adapting proximal equilibrium solvers especially Mirror Prox (and its NashMD variant)-to large language model fine-tuning, highlighting their fast, stable convergence via extragradient updates. We also discuss practical techniques for implementing approximate Mirror Prox efficiently at scale and clarify connections to RLHF/PPO and DPO-based fine-tuning.

Bio:

Michal is the Founding Researcher at Isara Labs, tenured researcher at Inria, and the lecturer at the MVA master of ENS Paris-Saclay. Michal is primarily interested in designing algorithms that would require as little human supervision as possible. That is why he is working on methods and settings that are able to deal with minimal feedback, such as deep reinforcement learning, bandit algorithms, self-supervised learning, or self play. Michal has recently worked on representation learning, word models and deep (reinforcement) learning algorithms that have some theoretical underpinning. In the past he has also worked on sequential algorithms with structured decisions where exploiting the structure leads to provably faster learning. Michal is now working on large large models (LMMs), in particular providing algorithmic solutions for their scalable fine-tuning and alignment. He received his Ph.D. in 2011 from the University of Pittsburgh, before getting a tenure at Inria in 2012 and starting Google DeepMind Paris in 2018 with Rémi Munos, In 2024, he became the principal Llama scientist at Meta, building online reinforcement learning stack and research for Llama 3.

Manifold Gaussian Mixture Models

Vladimir Spokoiny
HSE University, WIAS Berlin

The talk describes a new approach to modeling and estimating a high-dimensional distribution using the ideas of Gaussian Mixture Models (GMM) and manifold learning. The manifold hypothesis assumes that the GMM structure is preserved after projecting the data onto a low-dimensional

manifold. The procedure estimates the manifold projections in the original space and recovers the underlying GMM structure in a data-driven way using anisotropic smoothing. Numerical results show a state-of-the-art performance of the method. We also show that the approach ensures optimal identification of the Gaussian components in terms of the effective manifold dimension in place of the ambient data dimension.

Bio:

PhD in math from Lomonosov Moscow State University, Moscow, 1988. From 2000 to 2025: head of research group at WIAS Berlin and Prof. of Math. at HU Berlin. SInce 2025 - scientific head of TFAIM, HSE University.

Balancing Intelligence and Efficiency in Large Language Models

Bingyi Jing(荆炳义)
The Chinese University of Hong Kong, Shenzhen

Large language models can achieve impressive performance through extended reasoning, but often at the cost of increased computation and latency. This talk presents a statistically principled framework for adaptive reasoning that allocates computational effort according to task difficulty and uncertainty. By combining uncertainty-aware routing, off-policy evaluation, and sequential inference, the framework determines when expensive reasoning is necessary while providing rigorous guarantees on performance degradation. Theoretical results establish distribution-free reliability guarantees, and experiments on challenging reasoning benchmarks demonstrate substantial reductions in reasoning cost without sacrificing accuracy. This work illustrates how statistical methodology can help build AI systems that are both powerful and efficient.

Bio:

Bing-Yi Jing is Duan Family Presidential University Chair Professor at the School of Artificial Intelligence, The Chinese University of Hong Kong, Shenzhen, and Professor at the Shenzhen Loop Area Institute. He is a recipient of the National Natural Science Award (Second Class) and the Ministry of Education Natural Science Award for Higher Education Institutions (Second Class) and serves as a Changjiang Chair Professor appointed by the Ministry of Education of China. He is an ASA Fellow, an IMS Fellow. He has held associate editorships for several leading international academic journals.

Professor Jing's research interests span artificial intelligence, data science, econometrics, network data analysis, bioinformatics, probability, and statistics. He has published more than 140 papers in leading journals and conferences, including AoS, JRSS-B, JASA, Biometrika, AoP, JoE, JMLR, NeurIPS, ICLR, ICML, and CVPR. He also has extensive experience collaborating with industry and has received the Huawei Spark Award and the Huawei Outstanding Collaborative Achievement Award.

Flow Matching from Viewpoint of Proximal Operator

Kenji Fukumizu
The Institute of Statistical Mathematics

We reformulate Optimal Transport Conditional Flow Matching (OT-CFM), showing that it admits an

exact proximal form via an extended Brenier potential, without assuming that the target distribution has a density. More precisely, the mapping to recover the target point, or the "denoiser", is expressed by a proximal operator, which yields an explicit proximal expression of the vector field. Using the explicit proximal form with convex analysis, we analyze the stability of the dynamics under the manifold hypothesis, where the target distribution is supported on a low-dimensional manifold. We prove that the manifold structure is stable by perturbation to the dynamics: the dynamics does not expand or shrink exponentially along manifold directions. We also propose a new algorithm of few-step generation by learning the denoiser directly, and show its preliminary results.

Bio:

Kenji Fukumizu is a Professor at The Institute of Statistical Mathematics, where he also serves as Director of the Research Center for Statistical Machine Learning. He received the B.S. and Ph.D. degrees from Kyoto University in 1989 and 1996, respectively. He joined the Research and Development Center, Ricoh Co., Ltd in 1989, worked as a Researcher at the Institute of Physical and Chemical Research (RIKEN) from 1998, and became an Associate Professor at The Institute of Statistical Mathematics in 2000. He was a Visiting Scholar at the Department of Statistics, University of California, Berkeley in 2002-2003, and a Humboldt Fellow at Max Planck Institute for Biological Cybernetics in 2006--2007. He has been serving as Technical Advisor at Preferred Networks, Inc. since 2018, and at Fujitsu since 2025. He has served as an organizing committee member for many conferences, including Program Co-Chair of AISTATS 2021, and Senior Area Chair on the Program Committees of NeurIPS and ICML. He has also served as Chief Editor of the Annals of the Institute of Statistical Mathematics from 2011 to 2017 and currently serves as an Action Editor for the Journal of Machine Learning Research. He received the Awards for Science and Technology (Research Category) of the Commendation for Science and Technology by MEXT in 2026.

Stochastic optimal control approach to generative modelling and Schrödinger potential estimation

Nikita Puchkin
HSE University

Stochastic optimal control problem with a final constraint provides a natural way to construct a Schrödinger bridge between two distributions, making it well-suited for generative modelling. In this problem, the optimal control can be expressed through the Schrödinger potential,which depends on the target distribution --- typically unknown in practice. We address the problem of estimating this potential from finite samples. Focusing on estimators that minimize the empirical Kullback Leibler (KL) divergence, we study their generalization abilities. Despite the loss function’s unusual structure, we show that it exhibits favourable geometric properties under mild assumptions that hold for a broad class of target distributions. Finally, we derive non-asymptotic, high-probability upper bounds for the potential estimation accuracy, measured in terms of excess KL-risk.

Bio:

N. Puchkin received his B.Sc. and M.Sc. (Hons.) degrees in applied mathematics and physics from the Moscow Institute of Physics and Technology in 2016 and 2018, respectively, and an additional M.Sc. (Hons.) degree in mathematics and computer science from the Skolkovo Institute of Science and Technology in 2018. In 2023, he obtained a Ph.D. in mathematics from HSE University under the supervision of Prof. Dr. V. Spokoiny. He was awarded the 2020 Young Russian Mathematics Award and the 2024 Junior Leader Grant (mathematics). Since September 2024, he has been leading

the Laboratory for Theoretical Foundations of AI Models at HSE University.

Modulated learning for private and distributed regression with just a single sample per client device

Praneeth Vepakomma
MIT, MBZUAI

This work focuses on the question of learning from a large number of devices with each device holding only a single sample of data. Several real-world applications exist to this one sample per client setup up including learning from fitness trackers, data/app usage aggregators, body-worn sensing devices, and daily event monitors to name a few. When a client has only one sample, the standard federated learning paradigm breaks down as a local update based on that single point is far from being useful, especially in the earlier rounds for estimation of the model coefficients. This utility is further weakened by the privacy-inducing noise applied at every round. This work caters to this problem to enable such clients to collaboratively contribute to effectively learn a global model without leaking the privacy of their data. The proposed approach injects a single, carefully calibrated noisy perturbation to transform the sample at each client, followed by a post-processed representation which is shared with the server. These representations aggregated at the server are processed to obtain an unbiased gradient update that in expectation matches the non-privatecentralized gradient while preserving data privacy. This approach is different than traditional private federated learning, where the communication payloads involve model coefficients as opposed to privately transformed data samples. This method enables devices with extremely limited data to collaborate and learn accurate, privacy-preserving models without requiring large local datasets or sacrificing individual privacy.

Next-Latent Prediction Transformers Learn Compact World Models

Alex Lamb(杨丁怀)
Tsinghua University

Transformers replace recurrence with a memory that grows with sequence length and self-attention that enables ad-hoc look ups over past tokens. Consequently, they lack an inherent incentive to compress history into compact latent states with consistent transition rules. This often leads to learning solutions that generalize poorly. We introduce Next-Latent Prediction (NextLat), which extends standard next-token training with self-supervised predictions in the latent space. Specifically, NextLat trains a transformer to learn latent representations that are predictive of its next latent state given the next output token. Theoretically, we show that these latents provably converge to belief states, compressed information of the history necessary to predict the future. This simple auxiliary objective also injects a recurrent inductive bias into transformers, while leaving their architecture, parallel training, and inference unchanged. NextLat effectively encourages the transformer to form compact internal world models with its own belief states and transition dynamics -- a crucial property absent in standard next-token prediction transformers. Empirically, across benchmarks targeting core sequence modeling competencies -- world modeling, reasoning, planning, and language modeling -- NextLat demonstrates significant gains over standard next-token training in downstream accuracy, representation compression, and lookahead planning. NextLat stands as a simple and efficient paradigm for shaping transformer representations toward stronger generalization.

Bio:

Alex Lamb is an assistant professor at Tsinghua University AI College and a former senior researcher in the AI Frontiers group at Microsoft. He completed his PhD under Yoshua Bengio and has worked on deep learning, generative models, reinforcement learning, and sequence models. Alex Lamb received his Bachelor's degree in Computer Science and Applied Math and Statistics from Johns Hopkins University in 2012. He also worked on deep learning for classical Japanese document recognition as well as demand forecasting systems at Amazon.

Optimal in-context adaptivity and distributional robustness of transformers

Tengyao Wang(王藤耀)
London School of Economics

We study in-context learning for Transformer models trained on regression tasks with varying levels of difficulty. By establishing a new connection between in-context learning and nonparametric Bayesian regression, we prove for the first time that pretrained Transformers can adapt to the difficulty of a new task at test time and achieve faster convergence rates on smoother regression problems, after being trained on mixtures of tasks with heterogeneous complexity. Our results also show that this adaptation remains robust under distribution shift.

Bio:

Tengyao Wang is Professor in Statistics at the London School of Economics and Political Science. He received his PhD from the University of Cambridge. His research focuses on high-dimensional statistics and statistical machine learning, with particular interests in robust inference, adaptive learning, and change detection under complex nuisance variation. He has published widely in leading statistics journals and has received the Royal Statistical Society Guy Medal in Bronze and the RSS Research Prize.

Pointwise Generalization in Deep Neural Networks

Yunbei Xu(许云贝)
National University of Singapore

We address the fundamental question of why deep neural networks generalize by establishing a pointwise generalization theory for fully connected networks. This framework resolves long-standing barriers to characterizing the rich nonlinear feature-learning regime and builds a new statistical foundation for representation learning. For each trained model, we characterize the hypothesis via a pointwise Riemannian Dimension, derived from the eigenvalues of the learned feature representations across layers. This establishes a principled framework for deriving hypothesis-dependent, representation-aware generalization bounds. These bounds offer a systematic upgrade over approaches based on model size, products of norms, and infinite-width linearizations, yielding guarantees that are orders of magnitude tighter in both theory and experiment. Analytically, we identify the structural properties and mathematical principles that explain the tractability of deep networks. Empirically, the pointwise Riemannian Dimension exhibits substantial feature compression, decreases with increased over-parameterization, and captures the implicit bias of optimizers. Taken together, our results indicate that deep networks are mathematically tractable in

practical regimes and that their generalization is sharply explained by pointwise, feature-spectrum-aware complexity.

Bio:

Prof. Yunbei Xu is an Assistant Professor at the National University of Singapore. He received a B.S. in Pure Mathematics from Peking University and a Ph.D. in Decision Sciences from Columbia Business School, and completed postdoctoral training in computing at MIT. His work has received the ICML Outstanding Paper Award and first place in the INFORMS Student Paper Competition. His recent research is organized around two related themes: generalization in deep learning, with a focus on pointwise generalization and feature compression; and information-theoretic questions in reinforcement learning, including Bellman-sufficient information complexity and representation axioms.

Generalization bounds for Diffusion Models and Beyond

Konstantin Yakovlev
HSE University

Denoising Score Matching (DSM) is a cornerstone of diffusion-based generative models, yet establishing its generalization bounds typically reveals a severe curse of dimensionality. In this work, we analyze the statistical efficiency of DSM under a relaxed manifold assumption, where the data distribution is viewed as a perturbation of a low-dimensional Hölder-smooth manifold. We demonstrate that the true score function can be approximated by ReLU neural networks with a complexity scaling primarily with the intrinsic dimension rather than the ambient dimension. Furthermore, we extend our analysis to the estimation of the score Jacobian matrix and Implicit Score Matching (ISM). By introducing a novel application of the weighted Gagliardo-Nirenberg inequality, we prove that learning the Jacobian matrix is as efficient as learning the score itself. Our results show that ISM achieves the same sharp convergence rates as DSM, providing a rigorous theoretical foundation for the effectiveness of score-based models in high-dimensional spaces.

Bio:

Konstantin Yakovlev is a Junior Research Fellow at HSE University. He earned his Bachelor's and Master's degrees from Moscow Institute of Physics and Technology. His research interests lie at the intersection of nonparametric statistics and generative AI. His work has been featured in top-tier venues such as COLT, NeurIPS, ACL, EMNLP, and SIGIR.

Foundations of Diffusion Generative Models

Gen Li(李根)
The Chinese University of Hong Kong

Diffusion models have emerged as a powerful paradigm for generative modeling, enabling high-fidelity sampling by reversing a stochastic corruption process. Despite their empirical success, diffusion models face fundamental challenges in practice, especially in data-limited and compute-constrained regimes. This necessitates understanding and improving the sample complexity and computational efficiency of diffusion models. This lecture provides a unified mathematical

framework for diffusion models (as well as flow matching), covering important algorithmic developments and theoretical foundations. A central goal is to characterize the fundamental principles governing diffusion-based generation and to highlight computationally efficient algorithms. The lecture aims to equip attendees with a principled understanding of diffusion models and a modern toolbox for analyzing and designing faster and more reliable diffusion-based methodologies, along with a synthesis of open problems and future research directions.

Bio:

Gen Li is currently an assistant professor in the Department of Statistics and Data Science at the Chinese University of Hong Kong. His research interests include diffusion based generative model, reinforcement learning, high-dimensional statistics, machine learning

Diffeomorphic Markov Chain Monte Carlo: fast mixing for heavy-tailed distributions

Miha Bresar
The Chinese University of Hong Kong, Shenzhen

We introduce a new class of uniformly ergodic MCMC samplers and provide fast non-asymptotic mixing guarantees for target distributions on R^d with arbitrarily heavy polynomial tails. The novel samplers solve a well-known problem for state-of-the-art MCMC algorithms, which struggle with the combination of unbounded high-dimensional state space and vanishing gradients.

Our Diffeomorphic Contraction Sampler (DCS) transforms target π onto a Euclidean ball $B(R)$ and then samples from a transformed density on the convex set $B(R)$ via algorithms such as Ball Walk, Hit-and-Run and others. The radial diffeomorphic contraction guarantees a bounded density on $B(R)$, implying uniform ergodicity for *all* targets with a finite polynomial moment. Non-asymptotic bounds for DCS require stronger assumptions such as log-concavity of the density on $B(R)$. In practice, this is achieved via a preconditioned automorphism of the ball $B(R)$, tuned by Variational Inference.

Numerical simulation tests demonstrate that the DCS significantly outperforms the No-U-Turns sampler on high-dimensional heavy-tailed targets arising as real-world posteriors in PosteriorDB benchmark. DCS also numerically outperforms recently developed spherical projection samplers for heavy-tailed target distributions. The talk is based on recent joint work with Aleks Mijatovic (Warwick).

Bio:

Miha Bresar is an Assistant Professor in the School of Data Science at the Chinese University of Hong Kong, Shenzhen. He earned his PhD in Statistics from the University of Warwick in 2023, where his doctoral work was recognized with the G-Research Prize. His research interests lie at the intersection of probability theory and stochastic algorithms.

Stochastic Approximation Schemes with Decision-Dependent Samples: The Case of Performative Prediction

Hoi-To Wai(韋凱滔)
The Chinese University of Hong Kong

In recent years, stochastic approximation (SA) schemes have garnered renewed interest in the dynamic environment setting where streaming data is not independent and identically distributed, but rather is adaptive and/or decision-dependent. This presentation focuses on SA schemes applied to the performative prediction problem(s). We will motivate the problem through an illustrative example of strategic classification and demonstrate that a natural implementation of “SGD" with “greedy deployment” yields an SA scheme with non-gradient mean-fields. We will present recent results on the convergence of such SA scheme(s) under both convex and non-convex settings, as well as in stateful and non-stateful agent environments.

Bio:

Hoi-To Wai received his PhD degree from Arizona State University (ASU) in Electrical Engineering in Fall 2017, B. Eng. (with First Class Honor) and M. Phil. degrees in Electronic Engineering from The Chinese University of Hong Kong (CUHK) in 2010 and 2012, respectively. He is an Associate Professor in the Department of Systems Engineering & Engineering Management at CUHK. He has held research positions at ASU, UC Davis, Telecom ParisTech, Ecole Polytechnique, MIT. He is also an Associate Editor for the IEEE Transactions on Signal Processing and Elsevier’s Signal Processing. Hoi-To’s research interests are in the broad area of signal processing, machine learning and stochastic optimization. His dissertation has received the 2017’s Dean’s Dissertation Award from the Ira A. Fulton Schools of Engineering of ASU and he is a recipient of Best Student Paper Awards at ICASSP 2018, SAM 2024 (as a co-author), ICASSP 2025 (as a co-author).

Statistical analysis of Inverse Entropy-regularized Reinforcement Learning

Denis Belomestny
HSE University, Duisburg Essen University

Inverse reinforcement learning (IRL) aims to infer the reward function that explains expert behavior observed through trajectories of state--action pairs. A long-standing difficulty in classical IRL is the non-uniqueness of the recovered reward: many reward functions can induce the same optimal policy, rendering the inverse problem ill-posed. In this paper, we develop a statistical framework for *Inverse Entropy-regularized RL* that resolves this ambiguity by combining entropy regularization with a least-squares reconstruction of the reward from the soft Bellman residual. This combination yields a unique and well-defined so-called least-squares reward consistent with the expert policy. We model the expert demonstrations as a Markov chain with the invariant distribution defined by an unknown expert policy π^* and estimate the policy by a penalized maximum-likelihood procedure over a class of conditional distributions on the action space. We establish high-probability bounds for the excess KL--divergence between the estimated policy and the expert policy, accounting for statistical complexity through covering numbers of the policy class. These results lead to non-asymptotic minimax optimal convergence rates for the least-squares reward function, revealing the interplay between smoothing (entropy regularization), model complexity, and sample size. Our analysis bridges the gap between behavior cloning, IRL and modern statistical learning theory.

Bio:

Denis Belomestny is a W3 Professor of Applied Probability at Faculty of Mathematics in Duisburg-Essen University. His research interests lie in Mathematical Statistics, Reinforcement Learning and Optimal Stochastic Control with applications in Finance and Engeneering.

Wasserstein Policy Learning for Distributional Outcomes

Zhiheng Zhang(张智恒)
Shanghai University of Finance and Economics

Offline policy learning has received growing attention in causal inference. The primary objective is to learn a policy (individualized treatment rule) as a mapping from covariates to treatment that maximizes the empirical welfare defined as the mean of scalar-valued potential outcome. In this paper, we study offline policy learning with distribution-valued outcomes, where each potential outcome is a probability measure on $\mathbb{R}$ and the reward is defined through a utility functional applied to the Wasserstein barycenter of induced outcome distributions.

We establish statistical guarantees for the policy learning framework based on both Inverse Probability Weighting (IPW) and Doubly Robust (DR) estimators. By handling the challenging uniform deviation over the product of the combinatorial policy class and the infinite-dimensional quantile domain, we prove that the finite-sample regret scales as $\mathcal{O}(\sqrt{\text{N-dim}(\Pi)/N})$. This rate matches the classical scalar-outcome bound, revealing that the functional complexity of the distributional outcome space does not compromise the statistical efficiency. Moreover, we provide a matching minimax lower bound to establish the optimality of our proposed estimators.

Bio:

Zhiheng Zhang is a tenure-track Assistant Professor in the School of Statistics and Data Science at the Shanghai University of Finance and Economics (SUFE), starting from 2025.08. He is also affiliated with the Institute of Big Data Research. Previously, he received his Ph.D. from the Institute for Interdisciplinary Information Sciences, Tsinghua University, advised by Professor Yuhao Wang. He currently lead the CAUSAL lab at SUFE, focusing on developing principled methods for causal analysis by uncovering underlying structures and integrating them with modern learning and decision-making systems, with applications in ride-hailing and recommendation platform.

Statistical Certification for Majority Voting

Junpei Komiyama
Mohammed Bin Zayed University of Artificial Intelligence

Large language models often improve reasoning by sampling multiple outputs and aggregating their final answers, but precisely and efficiently controlling error levels remains a challenging task. In particular, deciding when to stop sampling remains difficult when the stopping rule is data-dependent and the set of possible answers is not known in advance. We study certification of the mode (majority) of the model's response distribution that provably controls the type-I error with adaptive sampling. The proposed method works without requiring prior knowledge of the answer category set.

Bio:

Professor Komiyama's research focuses on developing principled methods for machine learning, sequential decision-making, stochastic search, and the statistical foundations of reasoning in large language models. His most recent work includes best-arm identification in sequential decision-making settings and reliable chain-of-thought reasoning in LLMs, and anytime-valid certification of model outputs. He also works on mechanism design, fair allocation, and algorithmic bias, aiming to build algorithms with rigorous guarantees for learning and reasoning.

Functional Scaling Laws for LLM Pretraining

Lei Wu(吴磊)
Peking University

Scaling laws are among the most influential empirical regularities in LLM pretraining: final performance improves predictably as data, model size, and compute scale. However, classical scaling laws primarily describe the endpoint of training, leaving open the dynamical mechanism that produces these laws and how scaling behavior is shaped by hyperparameters.

In this talk, I will present Functional Scaling Laws (FSL), a framework that extends scaling laws from final-loss prediction to full loss-trajectory prediction. Starting from power-law kernel regression as a toy model, we derive stochastic training dynamics and reveal an intrinsic-time structure that unifies training curves across scales and protocols.

This functional perspective provides a dynamical theory of scaling laws. It explains how signal learning and noise forgetting jointly determine the loss dynamics, and shows how learning-rate and batch-size schedules shape scaling behavior through this balance. Experiments on LLM pretraining support FSL as a framework for interpreting and designing large-scale pretraining.

Bio:

Lei Wu is an Assistant Professor at the School of Mathematical Sciences and the Center for International Machine Learning Research at Peking University. His research focuses on the scientific foundations of deep learning. He received his B.S. in Mathematics from Nankai University in 2012, and his Ph.D. in Computational Mathematics from Peking University in 2018. From November 2018 to October 2021, he conducted postdoctoral research at Princeton University and the University of Pennsylvania. His work has been published in top conferences and journals such as NeurIPS, ICML, COLT, ICLR, Annals of Statistics, Journal of Machine Learning Research, IEEE Transactions on Information Theory.

Gaussian Approximation and Bootstrap Inference for Stochastic Gradient Methods and Reinforcement Learning

Marina Sheshukova
HSE University

Stochastic gradient and stochastic approximation algorithms are widely used in machine learning and reinforcement learning, but quantifying the uncertainty of their outputs remains challenging. In this talk, we discuss non-asymptotic Gaussian approximation results for Polyak–Ruppert averaged

iterates. We first consider stochastic gradient descent under independent sampling and establish non-asymptotic validity guarantees for a multiplier bootstrap procedure that constructs confidence sets without explicitly estimating the limiting covariance matrix.We then turn to linear stochastic approximation with Markovian noise, which covers a broad class of reinforcement learning algorithms. We establish Berry–Esseen bounds for the averaged iterates and provide non-asymptotic guarantees for a multiplier block bootstrap procedure for constructing confidence intervals under Markovian sampling.

Bio:

Marina Sheshukova is a Junior Research Fellow and a PhD student at HSE University, where she also received her Bachelor's and Master's degrees. Her research interests include statistical inference, stochastic approximation, and contextual bandits. Her work has been published at leading conferences, including AISTATS, NeurIPS, ICLR, and The ACM Web Conference.

Linear Algebra Problems and Algorithms in Muon-based Optimization Methods

Maxim Rakhuba
HSE University

In this talk, we will focus on muon-type optimizers that have recently gained popularity for training large language models. We will discuss linear algebra problems that have been forgotten and new algorithms for addressing them, driven by the demand for efficient implementations. We will also touch on new Muon modifications that are associated with efficient linear algebra structures.

Bio:

Maxim Rakhuba is an Associate Professor and the Head of the Laboratory for Matrix and Tensor Methods in Machine Learning at HSE University. He is a graduate of the Moscow Institute of Physics and Technology, earned his PhD from the Institute of Numerical Mathematics of the Russian Academy of Sciences, and completed a postdoctoral fellowship at ETH Zurich. His publication record includes top-tier machine learning venues (NeurIPS, ICLR, ICML, ECCV, CVPR, AISTATS, etc) and high-impact numerical journals (SISC, SINUM, NLAA, JCP, etc).

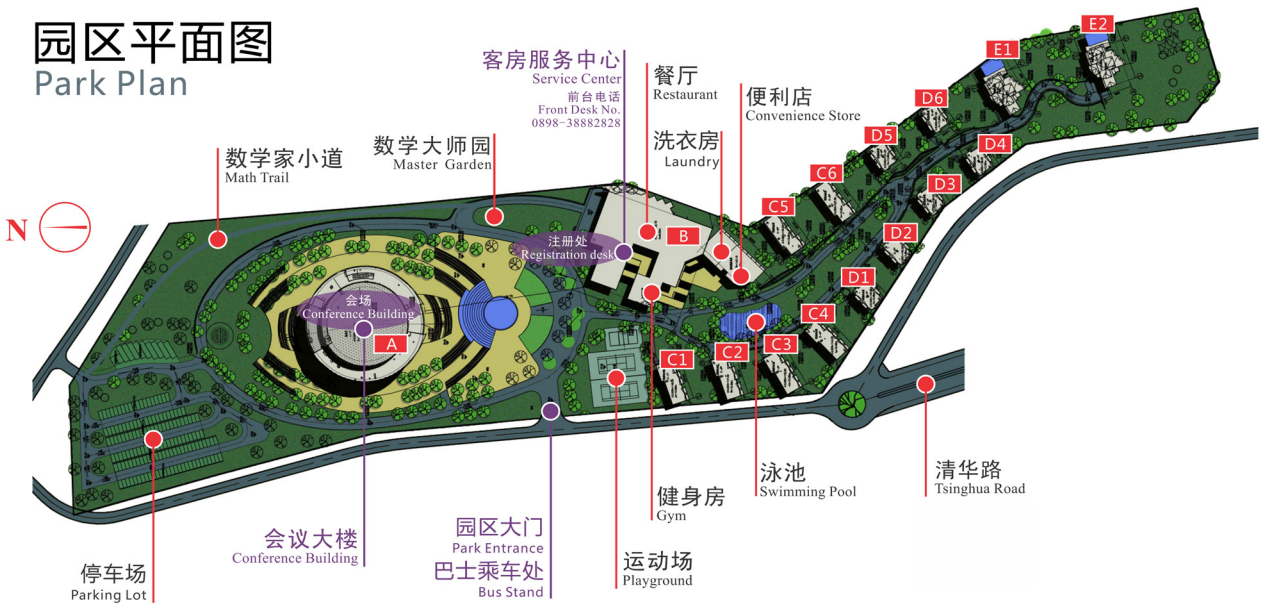

Welcome to TSIMF



The facilities of TSIMF are built on a 23-acre land surrounded by pristine environment at Phoenix Hill of Phoenix Township. The total square footage of all the facilities is over 29,000 square meter that includes state-of-the-art conference facilities (over 10,000 square meter) to hold many international workshops simultaneously, two reading rooms of library, a guest house (over 10,000 square meter) and the associated catering facilities, a large swimming pool, gym and sports court and other recreational facilities.

Management Center of Tsinghua Sanya International Forum is responsible for the construction, operation, management and service of TSIMF. The mission of TSIMF is to become a base for scientific innovations, and for nurturing of innovative human resource; through the interaction between leading mathematicians and core research groups in pure mathematics, applied mathematics, statistics, theoretical physics, applied physics, theoretical biology and other relating disciplines, TSIMF will provide a platform for exploring new directions, developing new methods, nurturing mathematical talents, and working to raise the level of mathematical research in China.

About Facilities



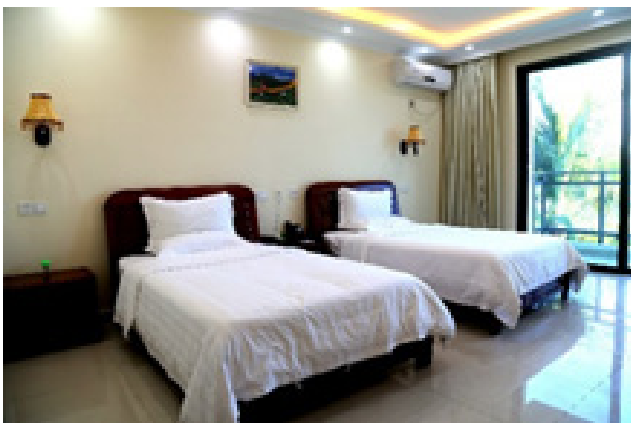
Registration

Conference booklets, room keys and name badges for all participants will be distributed at the front desk. Please take good care of your name badge. It is also your meal card and entrance ticket for all events.

Park-Wide WiFi Information

SSID:TSIMF-WiFi

Password:tsimf123



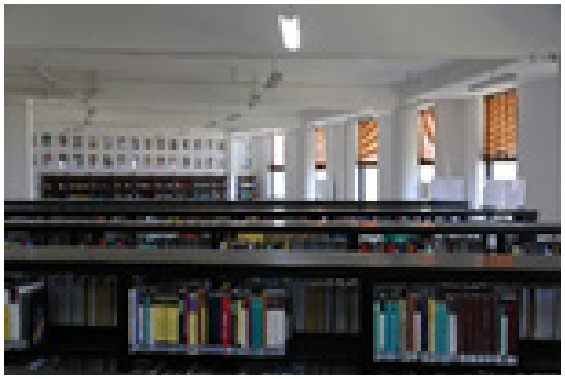
Family rooms are also equipped with kitchen and refrigerator.

Guest Room

All the rooms are equipped with: free Wi-Fi, TV, air conditioning and other utilities.



Library



Opening Hours: 09:00am-22:00pm

TSIMF library is available during the conference and can be accessed by using your room card. There is no need to sign out books but we ask that you kindly return any borrowed books to the book cart in library before your departure.



In order to give readers a better understanding of the contributions made by the Fields Medalists, the library of Tsinghua Sanya International Mathematics Forum (TSIMF) instituted the Special Collection of Fields Medalists as permanent collection of the library to serve the mathematical researchers and readers.

So far, there are 271 books from 49 authors in the Special Collection of Fields Medalists of TSIMF library. They are on display in room A220. The participants are welcome to visit.



Restaurant

All the meals are provided in the restaurant (Building B1) according to the time schedule.

Breakfast 07:30-08:30

Lunch 12:00-13:30

Dinner 17:30-19:00





Laundry

Opening Hours: 24 hours

The self-service laundry room is located in the Building(B1).



Gym

Opening Hours: 24 hours

The gym is located in the Building 1 (B1), opposite to the reception hall. The gym provides various fitness equipment, as well as pool tables, tennis tables etc.

Playground



Playground is located on the east of the central gate. There you can play basketball, tennis and badminton. Meanwhile, you can borrow table tennis, basketball, tennis balls and badminton at the reception desk.

Swimming Pool

Please enter the pool during the open hours, swimming attire and swim caps are required, if you feel unwell while swimming, please stop swimming immediately and get out of the pool. The depth of the pool is 1.2M-1.8M.

Opening Hours: 13:00-14:00 18:00-21:00



Free Shuttle Bus Service at TSIMF

We provide free shuttle bus for participants and you are always welcome to take our shuttle bus, all you need to do is wave your hands to stop the bus.

Destinations: Conference Building, Reception Room, Restaurant, Swimming Pool, Hotel etc.



Contact Information of Administration Staff

Location of Conference Affairs Office: Room 104, Building A

Tel: 0086-898-38263896

Conference Affairs :

Sarah 陈媛姗

Tel/Wechat:0086-130-2983-0780

Email: tsimf@tsinghua.edu.cn

Shouxi He 何守喜

Tel/Wechat:0086-186-8980-2225

Email: heshouxi@tsinghua.edu.cn

Location of Accommodation Affairs Office: Room 200, Building B1

Tel: 0086-898-38882828

Accommodation Manager: Ms. Li YE 叶莉

Tel/Wechat: 0086-139-7679-8300

Email: yel@tsinghua.edu.cn

*Reception duty hours: 7:00-23:00, chamber service please call: 0086-38882828 (exterior line) 80000 (internal line)

*Room maintainer night duty hours: 23:00-7:00, if you need maintenance services, please call: 0086-38263909 (exterior line) 30162 (internal line)

IT

Yuanhang Zhou 周远航

Tel/Wechat: 0086-133-6898-0169

Email: 13368980169@163.com

Director Assistant of TSIMF

Kai CUI 崔凯

Tel/Wechat: 0086- 136-1120-7077

Email :cuik@tsinghua.edu.cn

Director of TSIMF

Junpeng Zhu 朱俊鹏

Tel: 0086-136-1113-2615

Email: zjp@tsinghua.edu.cn

清华大学三亚国际论坛管理中心 | 三亚清华数学论坛管理中心
Tsinghua Sanya International Mathematics Forum (TSIMF)
0086-898-38882828
0086-898-38883896
0086-898-38883895
tsimf@tsinghua.edu.cn
572000
https://tsimf.tsinghua.edu.cn
海南省三亚市天涯区清华路 100 号 , 清华三亚国际数学论坛
No.100, Tsinghua Road, Tianya District, Sanya, Hainan, P. R. China.

Wechat

Service Quality
Satisfaction Survey

