Stefano Sarao Mannelli

Assistant Professor,
Data Science and AI, Computer Science and Engineering,
Chalmers University of Technology & Gothenburg University.
stefano.sarao@gmail.com
stefsmlab.github.io
scholar.google.com/citations?user=Kq272_MAAAAJ

Research Experience

- Sept. 2024 – Present: WASP-AI **Assistant Professor Tenure Track** in the Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, Sweden. I use statistical physics to study how neural networks learn. I ask which choices about data and training cause a model to make more errors for one group than another, and when changing the training procedure can reduce that gap.
- Feb. 2025 – Present: **Visiting Lecturer** in the School of Computer Science and Applied Mathematics, University of the Witwatersrand, South Africa. I designed and teach Modern Theories of Machine Learning and co-supervise Gabriel Drew's master's research. With Devon Jarvis and colleagues, I study what happens to poorly represented languages when models train on text generated by earlier models.
- Jun. 2021 – Aug. 2024: **Research Fellow → Senior Research Fellow** in Gatsby Computational Neuroscience Unit and Sainsbury Wellcome Centre, University College London, UK. With Andrew Saxe and colleagues, I developed models of how animals learn complex tasks in stages and how neural networks forget earlier tasks as they learn new ones.
- Oct. 2020 – Jun. 2021: **Research Fellow** in the Department of Experimental Psychology, University of Oxford, UK. I began working with Andrew Saxe on the mathematics of curriculum learning: when does seeing easier examples first help a network learn?
- Mar. 2020 - Apr. 2020: **Visiting researcher** with **Eric Vanden-Eijnden** at NYU, USA.
- Jan. 2019 - Mar. 2019: **Visiting researcher** at KITP, UCSB, U.S.A.
- Feb. 2018: **Visiting researcher** at Duke University, U.S.A.
- Mar. 2017 - Jul. 2017: **Research intern** with **Lenka Zdeborová** at IPhT, CEA, France.
- Jan. 2016 - Apr. 2016: **Research intern** with **Nicolas Vayatis** at CMLA, ENS Cachan, France.

Education

- Oct. 2017 – Oct. 2020: **Ph.D. in Theoretical Physics** at IPhT, CEA, France - Supervisor: **Lenka Zdeborová**;
- Sep. 2016 - Oct. 2017: **M.Sc. in Electronic Engineering** at Politecnico di Torino, Italy, grade: *110/110 cum laude*;
- Sep. 2014 - Jul. 2016: **M.Sc. in Physics of Complex Systems** at Politecnico di Torino-SISSA, Italy, grade: *110/110 cum laude*;
- Sep. 2014 - Jul. 2016: **M2, Physique Théorique** at Paris Diderot, UPMC, and ENS Cachan, France;
- Sep. 2014 - Jul. 2016: **Alta Scuola Politecnica Diploma** at Politecnico di Torino, Politecnico di Milano, Italy;
- Sep. 2014 - Jun. 2017: **M.Sc. in Engineering Physics** at Politecnico di Milano, Italy, grade: *110/110 cum laude*;
- Sep. 2011 - Jul. 2014: **B.Sc. in Mathematics for Engineering** at Politecnico di Torino, Italy, grade: *110/110 cum laude*.

Organisation of scientific events

Aug. 2027 **School on Analytical Connectionism**. School Director. Boston, USA.

Jun.-Jul. 2027 **KITP Program on Statistical Physics & Neurobiology of Learning in High-Level Cognition**. Organizer (forthcoming). Santa Barbara, USA.

Jul. 2027 **Conference on the Neuroscientific and Psychological Theories of Development in the Deep Learning Era**. Organizer (forthcoming). Santa Barbara, USA.

27–28 Oct. 2026 **CHAIR Structured Learning Workshop**. Co-organiser. Gothenburg, Sweden.

Aug. 2026 **School on Analytical Connectionism**. School Director. Gothenburg, Sweden. Secured 500,000 SEK.

Jun. 2026 **Workshop on the Mathematical Foundations of AI**. Organizer. Gothenburg, Sweden. Secured 165,000 SEK.

Mar. 2026 **Workshop on LLMs and Responsible AI**. Organizer. Gothenburg, Sweden.

Dec. 2025 **EurIPS Workshop - Unifying Perspectives on Learning Biases**. Organizer. Copenhagen, Denmark.

Aug. 2025 **School on Analytical Connectionism**. Organizer. London, UK. Secured £25K funding.

May 2025 **Workshop in Advancements in High-Dimensional Methods for Machine Learning**. Organizer. Gothenburg, Sweden. Secured 135,500 SEK.

Apr. 2025 **Introduction to Statistical Physics for ML Theory**. Organizer. Gothenburg, Sweden.

Aug. 2024 **School on Analytical Connectionism**. Organizer. New York, USA. Secured $152K funding.

Mar. 2024 **Workshop on Bridging Analytical and Experimental Insights into Representational Change (CoSyNe)**. Organizer. Lisbon, Portugal.

Oct. 2023 **Workshop on Analytical Approaches for Neural Network Dynamics**. Organizer. Paris, France. Secured €10K funding.

Aug.–Sep. 2023 **School and Workshop on Analytical Connectionism**. Organizer. London, UK.

Jun. 2022 **Workshop on Communication Across Communities in Machine Learning Research and Practice (FAccT 2022)**. Organizer. Seoul, South Korea.

May 2021 **Workshop on Science and Engineering of Deep Learning (ICLR 2021)**. Organizer. Virtual.

Ongoing Projects

- **Bias Generation and Amplification**: I study how group frequency and feature geometry affect learning over time. I track when classification errors diverge between groups and when a generative model begins to memorise examples from each group. I use these results to test whether changing how often each group is sampled can narrow a performance gap. (*AI Alignment Grant, 2026*)
- **Connecting Dynamics and Landscapes in High Dimensions:** I study how much the geometry of a loss landscape can tell us about the solution a learning algorithm will find. I compare the dynamics of gradient descent with the curvature near saddle points. In a model of signal recovery, this explains how the algorithm can recover the signal even when the landscape contains exponentially many poor local minima. (*WASP Starting Pkg 2025-*)
- **Bridging Biological and Artificial Neural Networks:** I study how people, animals and artificial networks learn a sequence of related tasks. I model how networks reuse what they learned on an earlier task to explain when knowledge transfers and when it is forgotten. By giving people and networks the same tasks in different orders, I test which aspects of human learning these models capture. (*WASP Starting Pkg 2025-*)

Research supervision

Flavio Nicoletti. Postdoctoral researcher, Chalmers.

Chenxiao Ma. PhD student, Chalmers.

Loek van Rossem. PhD student, UCL and Chalmers research groups. Co-supervised with Andrew Saxe.

Gabriel Ocana Santero. Doctoral research supervision, UCL (2024).

Gabriel Drew. Master's research co-supervision, Wits (2026).

Bingyu Zhu. Master's research, Chalmers (2026).

Anton O'Nils. Master's research, Chalmers and AstraZeneca (2026). Potency prediction from dose–response curves.

Elínborg Ása Ásbergsdóttir. Master's research, Chalmers (2025).

Andres Felipe Otero Salamanca. Master's research, Chalmers (2025).

Tobias Olsson. Master's research, Chalmers (2025).

Nicolas Johansson. Master's research, Chalmers (2025).

Jie Huang. Master's research, Chalmers (2025). Loss landscapes of two-layer ReLU networks.

Anchit Jain. Master's research, UCL (2024). How group disparities evolve during training.

Research mentoring

Mar.–Dec. 2023 **Jin Lee**. Doctoral researcher, University College London.

Apr.–Jun. 2023 **Léo Touzo**. Master's student, École normale supérieure.

Sept. 2022–Aug. 2023 **Nishil Patel**. Research assistant, University College London.

Feb.–Jul. 2022 **Gaspard Abel**. Master's student, École normale supérieure.

Teaching

Spring 2026 **Introduction to Data Science and AI**. Chalmers University of Technology and University of Gothenburg. Course examiner.

Feb. 2026 **Introduction to Modern Theories of Machine Learning**. University of the Witwatersrand. Designed and taught a master's course on the foundations of learning systems.

Autumn 2024 (LP2) **The Computer Scientist in Society**. Chalmers University of Technology and University of Gothenburg. Taught student presentation sessions.

Apr. 2024 **Statistical Physics of Learning**. University of the Witwatersrand. Short lecture series.

Jun. 2022 **Statistical Physics of Deep Learning**. Summer school, Como, Italy. Teaching assistant, 8 hours.

Mar.–May 2019 **Statistical Physics**. CentraleSupélec, France. Teaching assistant with Grégory Schehr, 15 hours.

Mar.–Jun. 2018 **Statistical Physics**. CentraleSupélec, France. Teaching assistant with Grégory Schehr, 20 hours.

Invited presentations and panels

3–4 Dec. 2026 **3rd Nordic Biomathematics Meeting**. Uppsala University, Sweden. Invited plenary, "Curriculum learning in humans and neural networks" (forthcoming).

17 Nov. 2026 **PIAMI Onboarding Workshop**. London, UK. Invited panellist, "Limits of theory and the methodological frontier" (forthcoming).

29 Oct. 2026 **AI Foundations panel, Chalmers AI Days**. Gothenburg, Sweden. Moderator.

Jun. 2026 **New Trends in Non-equilibrium Dynamics**. Kavli Institute for Theoretical Physics, Santa Barbara, USA. Talk on the loss landscape of two-layer ReLU networks.

May 2026 **BrainNet 2026**. Stockholm, Sweden. Talk on curriculum learning.

Mar. 2026 **MIND Institute Seminar**. University of the Witwatersrand, Johannesburg, South Africa. Group research goals and recent results.

Nov. 2025 **MLLS Seminar**. Copenhagen, Denmark. Talk on bias dynamics.

Aug. 2025 **Statistical Physics & Machine Learning: Moving Forward**. Cargese, France.

Jul. 2025 **6th Youth in High-Dimensions: Recent Progress in Machine Learning, High-Dimensional Statistics and Inference & How creative is Generative AI? Perspectives from Science and Philosophy**. Trieste, Italy.

Apr. 2025 **Workshop on Spurious Correlation and Shortcut Learning: Foundations and Solutions (ICLR 2025)**. Singapore.

Apr. 2025 **G-Cog Seminar Series**. Gothenburg, Sweden. Talk on curriculum learning and cognitive science.

Mar. 2025 **AI for active matter: From plankton to robots**. Gothenburg, Sweden.

Mar. 2025 **14th Nordic Workshop on Statistical Physics: Biological, Complex and Non-Equilibrium Systems**. Stockholm, Sweden.

Dec. 2024 **Fairness of AI in Medical Imaging**. Virtual.

Nov. 2024 **At the Crossroads of AI and Cognitive Science**. Gothenburg, Sweden.

Nov. 2024 **Informal Workshop on the Problem of Class Imbalance**. Paris, France.

Oct. 2024 **2024 CHAIR Structured Learning Workshop**. Gothenburg, Sweden.

Conference presentations and seminars

Jul. 2023 **Statistical Physics and ML Back Together Again**. Talk and poster, Cargèse, France.

2022 **International Conference on Machine Learning**. Talk and poster, Baltimore, USA.

Jun. 2022 **Youth in High Dimensions**. Posters, Trieste, Italy.

May 2022 **Tri-Centre meeting**. Talk, Hebrew University of Jerusalem, Israel.

Apr. 2022 **Bocconi University**. Seminar, Milan, Italy.

Mar. 2022 **Computational and Systems Neuroscience (COSYNE)**. Posters, Lisbon and Cascais, Portugal.

Nov. 2021 **King's College London**. Seminar, London, UK.

Jun. 2021 **Youth in High Dimensions**. Posters, online.

Dec. 2020 **NeurIPS**. Posters, online.

Sept. 2020 **Computational Phase Transitions**. Talk, Simons Institute, Berkeley, USA.

Mar. 2020 **City University of New York**. Seminar, New York, USA.

Feb. 2020 **École polytechnique fédérale de Lausanne**. Seminar, Lausanne, Switzerland.

Dec. 2019 **NeurIPS**. Talk and poster, Vancouver, Canada.

Jul. 2019 **Theoretical Advances in Deep Learning**. Talk, Türkiye.

Jun. 2019 **International Conference on Machine Learning**. Talk and poster, Long Beach, USA.

Jun. 2019 **Day on Methods for Dynamical Mean-Field Equations**. Talk, Paris, France.

Feb. 2019 **Machine Learning for Quantum Many-Body Physics**. Poster, KITP, Santa Barbara, USA.

Jan. 2019 **The Rough High-Dimensional Landscape Problem**. Poster, KITP, Santa Barbara, USA.

Sept. 2017 **Junior Conference on Data Science and Engineering**. Talk, France.

Grants and awards

Jun. 2026 **Research Visit Grant**. Helge Ax:son Johnsons stiftelse, 39,000 SEK.

May 2026 **Vetenskapsfond**. Wilhelm och Martina Lundgrens stiftelser, 50,000 SEK.

Jan. 2026 **AI Alignment Grant**. AISI, 969,811.36 GBP.

2025 **Top Reviewer Award**. NeurIPS.

Jan. 2025 **Academic Grant**. CM Lerici Foundation, 60,000 SEK.

Jul. 2024 **Travel Grant**. Guarantor of Brains, £1,000.

Feb. 2024 **Travel Grant**. G-Research, £1,400.

Nov. 2023 **UK–IT Trustworthy AI Exchange Programme**. The Alan Turing Institute, £4,500.

Mar. 2023 **SCGB Conference Awards**. Simons Foundation, £3,000.

Sept. 2017 **Ph.D. Scholarship**. CEA.

Feb. 2017 & Feb. 2016 **Mobility Grant "Tesi su proposta"**. Politecnico di Torino, €2,200 each.

Sept. 2011, 2012 & 2013 **Accommodation Scholarship "Borsa Talenti"**. Fondazione CEUR.

Sept. 2012 **3rd Prize – Individual Math Competition Alfa Class**. Fondazione CRT, €500.

Leadership and professional service

Founding team member and principal investigator. Principia.

2023–present **Co-founder and co-director**. Analytical Connectionism schools.

Research theme lead. CHAIR Mathematical Foundations of AI.

Vice-leader. WASP Mathematical Foundations of AI cluster.

Academic Year 2025–2026 **Organizer, Colloquium Series**. Department of Computer Science and Engineering, Chalmers University of Technology.

Academic Year 2022–2023 **Organizer, Neuroscience External Seminars**. Gatsby Computational Neuroscience Unit.

Academic Year 2021–2022 **Organizer, SaxeLab Seminars**.

2026 **Technical Program Committee member**. Conference on Cognitive Computational Neuroscience (CCN).

Action Editor. TMLR.

Area Chair. NeurIPS, ICLR, ICML.

Reviewer. NeurIPS, ICML, ICLR, SciPost, JSTAT, JPhysA, Physica A, PNAS, PRX, TMLR, PRE, EurIPS Workshops.

Publications and preprints

1. F. Kovačević, E. Cyffers, S. Sarao Mannelli and M. Mondelli. High-Dimensional Asymptotics and Dataset Selection for Private Transfer Learning. arXiv:2610.02578 (2026), preprint.

2. A. O'Nils, S. Sarao Mannelli, O. Engkvist, J. Sjölund and A. M. Westerlund. Potency Prediction via Probabilistic Modeling of Dose–Response Curves. ChemRxiv (2026), preprint. doi:10.26434/chemrxiv.15008694/v1.

3. F. Nicoletti, C. Ma, E. Ventura, L. Saglietti and S. Sarao Mannelli. The Interplay of Data Structure and Imbalance in the Learning Dynamics of Diffusion Models. NeurIPS 2026, forthcoming; arXiv:2605.06367.

4. D. Jarvis, R. Klein, B. Rosman, S. James and S. Sarao Mannelli. Position: the Stochastic Parrot in the Coal Mine. Model Collapse is a Threat to Low-Resource Communities. ICML 2026 (Spotlight); arXiv:2605.04127.

5. J. Huang, B. Loureiro and S. Sarao Mannelli. Sharp description of local minima in the loss landscape of high-dimensional two-layer ReLU neural networks. ICML 2026; arXiv:2604.09412.

6. K. J. Sandbrink, S. Sarao Mannelli and F. Krzakala. Thinking of Neural Networks Like a Physicist: The Statistical Physics of Machine Learning. Proceedings of the Analytical Connectionism Schools 2023-2024, PMLR 320:15-41 (2026).

7. Y. Strittmatter, S. Sarao Mannelli, M. Ruiz-Garcia, S. Musslick and M. W. H. Spitzer. Curriculum Learning in Humans and Neural Networks. Open Mind 10:739-753 (2026); earlier version in Proceedings of the Annual Meeting of the Cognitive Science Society 47 (2025).

8. S. Sarao Mannelli, F. Gerace, N. Rostamzadeh and L. Saglietti. Bias-inducing geometries: An exactly solvable data model with fairness implications. Physical Review E 112, 025304 (2025).

9. D. Jarvis, S. Lee, C. C. J. Dominé, A. M. Saxe and S. Sarao Mannelli. A Theory of Initialisation's Impact on Specialisation. ICLR 2025; Journal of Statistical Mechanics: Theory and Experiment 2025, 114001.

10. F. Mori, S. Sarao Mannelli and F. Mignacco. Optimal Protocols for Continual Learning via Statistical Physics and Control Theory. ICLR 2025; Journal of Statistical Mechanics: Theory and Experiment 2025, 084004.

11. N. Patel, S. Lee, S. Sarao Mannelli, S. Goldt and A. Saxe. RL Perceptron: Generalization Dynamics of Policy Learning in High Dimensions. Physical Review X 15, 021051 (2025).

12. F. Gerace, D. Doimo, S. Sarao Mannelli, L. Saglietti and A. Laio. How to choose the right transfer learning protocol? A qualitative analysis in a controlled set-up. Transactions on Machine Learning Research (2024).

13. Y. Li, R. Carrasco-Davis, S. Sarao Mannelli, Y. Strittmatter and S. Musslick. A Meta-Learning Framework for Rationalizing Cognitive Fatigue in Neural Systems. Proceedings of the Annual Meeting of the Cognitive Science Society 46 (2024).

14. A. Jain, R. Nobahari, A. Baratin and S. Sarao Mannelli. Bias in Motion: Theoretical Insights into the Dynamics of Bias in SGD Training. NeurIPS 2024.

15. S. Sarao Mannelli, Y. Ivashinka, A. Saxe and L. Saglietti. Tilting the Odds at the Lottery: the Interplay of Overparameterisation and Curricula in Neural Networks. ICML 2024.

16. J. H. Lee, S. Sarao Mannelli and A. Saxe. Why Do Animals Need Shaping? A Theory of Task Composition and Curriculum Learning. ICML 2024.

17. L. Saglietti, S. Sarao Mannelli and A. Saxe. An Analytical Theory of Curriculum Learning in Teacher-Student Networks. NeurIPS 2022; Journal of Statistical Mechanics: Theory and Experiment 2022, 114014.

18. S. Lee, S. Sarao Mannelli, C. Clopath, S. Goldt and A. Saxe. Maslow's Hammer in Catastrophic Forgetting: Node Re-Use vs. Node Activation. ICML 2022.

19. F. Gerace, L. Saglietti, S. Sarao Mannelli, A. Saxe and L. Zdeborová. Probing transfer learning with a model of synthetic correlated datasets. Machine Learning: Science and Technology 3, 015030 (2022).

20. S. Sarao Mannelli and P. Urbani. Analytical Study of Momentum-Based Acceleration Methods in Paradigmatic High-Dimensional Non-Convex Problems. NeurIPS 2021.

21. A. Baker, I. Biazzo, A. Braunstein, G. Catania, L. Dall'Asta, A. Ingrosso, F. Krzakala, F. Mazza, M. Mézard, A. P. Muntoni, M. Refinetti, S. Sarao Mannelli and L. Zdeborová. Epidemic mitigation by statistical inference from contact tracing data. Proceedings of the National Academy of Sciences (2021).

22. S. Sarao Mannelli, E. Vanden-Eijnden and L. Zdeborová. Optimization and Generalization of Shallow Neural Networks with Quadratic Activation Functions. NeurIPS 2020.

23. S. Sarao Mannelli, G. Biroli, C. Cammarota, F. Krzakala, P. Urbani and L. Zdeborová. Complex Dynamics in Simple Neural Networks: Understanding Gradient Flow in Phase Retrieval. NeurIPS 2020.

24. S. Sarao Mannelli and L. Zdeborová. Thresholds of descending algorithms in inference problems. Journal of Statistical Mechanics: Theory and Experiment 2020, 034004.

25. S. Sarao Mannelli, G. Biroli, C. Cammarota, F. Krzakala, P. Urbani and L. Zdeborová. Marvels and Pitfalls of the Langevin Algorithm in Noisy High-Dimensional Inference. Physical Review X 10, 011057 (2020).

26. S. Sarao Mannelli, G. Biroli, C. Cammarota, F. Krzakala and L. Zdeborová. Who is Afraid of Big Bad Minima? Analysis of gradient-flow in spiked matrix-tensor models. NeurIPS 2019 (Spotlight).

27. S. Sarao Mannelli, F. Krzakala, P. Urbani and L. Zdeborová. Passed & Spurious: Descent Algorithms and Local Minima in Spiked Matrix-Tensor Models. ICML 2019.