

Soufiane Hayou

Simons Institute for the Theory of Computing, UC Berkeley
✉ hayou@berkeley.edu
🌐 <https://www.soufianehayou.com/>

Academic Experience

Simons Institute, UC Berkeley

Researcher

Member of the Collaboration on the Theoretical Foundations of Deep Learning.
Developing the theory and practice of neural network scaling for both pretraining and finetuning.
(Co)Inventor of Stable ResNet, Depth- μ P and LoRA+.
Hosted by Bin Yu and Peter Bartlett.

Berkeley, United States

2023 – Current

National University of Singapore

PTA Assistant Professor of Mathematics

Courses taught (Graduate level):

- Principles of Machine Learning (DSA5105, 100+ students)
- Advanced Topics in Machine Learning (DSA5202)

Singapore

2021 – Current

University of Oxford

Class Tutor and Teaching Assistant

- Advanced simulation methods (Undergraduate level)
- Bayesian Inference and MCMC methods (Graduate level)

Oxford, United Kingdom

2018 – 2020

Education

University of Oxford

PhD in Statistics & Machine Learning. Advised by A. Doucet and J. Rousseau.

Thesis: Wide Deep Neural Networks. Worked on the theory of large scale neural networks, and derived actionable recipes for neural network scaling.

Oxford, United Kingdom

2017 – 2021

Pierre et Marie Curie University (Paris VI)

MSc. in Probability and Financial Mathematics with Distinction.

Courses included stochastic calculus, stochastic control, portfolio optimization (Markowitz's theory), point processes, high frequency trading, derivatives pricing.

Paris, France

2016 – 2017

Ecole Polytechnique

MSc. in Applied Mathematics and Engineering Diploma with Distinction.

Major courses: Markov chains, stochastic calculus and martingales, Fourier analysis, statistical analysis and model selection.

Minor courses: macroeconomics, microeconomics, sociology, political science.

Paris, France

2013 – 2017

Awards and Honours

- 2024: Gradient AI Fellowship
- 2023: Google Cloud TPU Credits Award
- 2023: Selected as a Rising Star in AI by KAUST
- 2022: Top Reviewer at NeurIPS
- 2022: SlowDNN Workshop Travel Grant
- 2021: Institute of Mathematical Science(NUS) Travel Award
- 2019: James Fund for Mathematics Travel Grant (One of two awarded campus-wide from over 100 applications)
- 2019: ICML 2019 Student Travel Award

- 2018: Natixis Best Master Thesis in Quantitative Finance (One of two awarded from over 200 submissions)
- 2017 - 2021: ESPRC Fellowship Grant
- 2017 - 2021: RCUK Fellowship Grant (St John's College)
- 2013 - 2017: Egide Grant (One of twenty awarded by the French Ministry of Foreign Affairs from over 1000 applications)
- 2011: Maroc Telecom Excellence Grant (awarded to top 1% students in Morocco)
- 2011: Moroccan Government Excellence Scholarship

Research

Research Interests

Main Research: I am currently working on the *theory* and *practice* of neural network scaling. I derive principled guidelines on how to scale width, depth, data, compute, hyperparameters etc, for both pretraining and finetuning of foundation models (LLMs, LVMs etc). Currently, I am focusing on width/depth parametrization for pretraining and finetuning. More precisely, based on infinite width/depth dynamics, my research characterizes the optimal choices of initialization and learning rate. Examples include Stable ResNet and Depth- μ P (Tensor Programs VI) for pretraining, and LoRA+ for finetuning large language models (LLMs).

Broad Research Interests: In addition to my main research, I am broadly interested in the following topics:

- Theory: Model and Data compression, Interplay between Neural Networks and Stochastic processes.
- Applications: Privacy in Deep Learning models, Machine Learning for Healthcare, Machine Learning for Finance.

Publications and Preprints

- ✦ *The Impact of Initialization on LoRA Finetuning Dynamics*. (submitted, 2024). Soufiane Hayou, Nikhil Gosh, Bin Yu.
- ✦ *LoRA+: Efficient Low Rank Adaptation of Large Models*. **International Conference on Machine Learning (ICML 2024)**. Soufiane Hayou*, Nikhil Gosh*, Bin Yu. (*equal contribution).
- ✦ \longleftrightarrow *WD(II): Commutative Width and Depth Scaling in Deep Neural Networks*. 2024. Accepted at **Journal of Machine Learning Research (JMLR)**. Soufiane Hayou.
- ✦ *How Bad is Training on Synthetic Data? A Statistical Analysis of Language Model Collapse*. **Conference on Language Modelling (COLM 2024)**. Mohamed El Amine Seddik, Swei-Wen Chen, Soufiane Hayou, Pierre Youssef, Merouane Debbah.
- ✦ *Tensor Programs VI: Feature Learning in Infinite Depth Neural Networks*. **International Conference on Learning Representations (ICLR 2024)**. Greg Yang, Dingli Yu, Chen Zhu, Soufiane Hayou.
- ✦ *Leave-one-out Distinguishability in Machine Learning*. **International Conference on Learning Representations (ICLR 2024)**. Jiayuan Ye, Anastasia Borovykh, Soufiane Hayou, Reza Shokri.
- ✦ \longleftrightarrow *WD(I): Width and Depth Limits commute in Residual Networks*. **International Conference on Machine Learning (ICML 2023)**. Soufiane Hayou, Greg Yang.
- ✦ *Data Pruning and Neural Scaling Laws: Fundamental Limitations of Score-based Algorithms*. **Transactions on Machine Learning Research (TMLR, 2023)**. Fadhel Ayed*, Soufiane Hayou*. (*equal contribution, alphabetical order).
- ✦ *On the Infinite-depth Limit of Finite-width Neural Networks*. **Transactions on Machine Learning Research (TMLR, 2022)**. Soufiane Hayou.

- ✧ *Feature Learning and Signal Propagation in Deep Neural Networks*. **International Conference on Machine Learning (ICML 2022, Spotlight presentation)**. Yizhang Lou, Chris Mingard, Yoonsoo Nam, Soufiane Hayou.
- ✧ *From Optimization Dynamics to Generalization Bounds via Lojasiewicz Gradient Inequality*. **Transactions on Machine Learning Research (TMLR, 2022)**. Fusheng Liu, Qianxiao Li, Soufiane Hayou, Haizhao Yang.
- ✧ *Regularization in ResNet with Stochastic Depth*. **Neural Information Processing Systems (NeurIPS 2021)**. Soufiane Hayou, Fadhel Ayed.
- ✧ *Probabilistic Fine-tuning of Pruning Masks and PAC-Bayes Self-bounded Learning*. (2021, arXiv). Soufiane Hayou, Bobby He, Gintare Karolina Dziugaite.
- ✧ *Robust Pruning at Initialization*. **International Conference on Learning Representations (ICLR 2021)**. Soufiane Hayou, Jean-Francois Ton, Arnaud Doucet, Yee Whye Teh.
- ✧ *Stable ResNet*. **The International Conference on Artificial Intelligence and Statistics (AISTATS 2021, Oral presentation)**. Soufiane Hayou*, Eugenio Clerico*, Bobby He*, George Deligiannidis, Arnaud Doucet, Judith Rousseau. (*equal contribution).
- ✧ *On the Impact of the Activation function on Deep Neural Networks Training*. **International Conference on Machine Learning (ICML 2019)**. Soufiane Hayou, Arnaud Doucet, Judith Rousseau.
- ✧ *Mean-field Behaviour of Neural Tangent Kernel for Deep Neural Networks*. (2020, submitted). Soufiane Hayou, Arnaud Doucet, Judith Rousseau.

→ Workshop papers:

- ✧ *Data Pruning and Neural Scaling Laws: Fundamental Limitations of Score-based Algorithms*. **ICML 2024 Workshop on Data-Centric Machine Learning Research**. Fadhel Ayed*, Soufiane Hayou*. (*equal contribution, alphabetical order). Full paper published at Transactions on Machine Learning Research (TMLR).
- ✧ *The curse of (non)convexity: The case of an Optimization-Inspired Data Pruning algorithm*. **NeurIPS 2022 ICBINB workshop**. Fadhel Ayed, Soufiane Hayou.
- ✧ *The Curse of Depth in Kernel Regime*. **NeurIPS 2021 ICBINB workshop (Spotlight)**. Also published as part of workshop volume at **Proceedings of Machine Learning Research (PMLR)**. Soufiane Hayou, Arnaud Doucet, Judith Rousseau.
- ✧ *Stochastic Pruning: Fine-Tuning, and PAC-Bayes Bound Optimization*. **NeurIPS 2021 Bayesian Deep Learning Workshop**. Soufiane Hayou, Bobby He, Gintare Karolina Dziugaite.

→ Notes and Technical Reports:

- ✧ *On the Connection Between Riemann Hypothesis and a Special Class of Neural Networks*. (2023, arXiv). Soufiane Hayou.
- ✧ *On the Overestimation of the Largest Eigenvalue of a Covariance Matrix*. (2017, Bloomberg). Soufiane Hayou.
- ✧ *Cleaning the Correlation Matrix with a Denoising AutoEncoder*. (2017, Bloomberg). Soufiane Hayou.

Academic Services

- Reviewing for conferences: NeurIPS, ICML, ICLR, AISTATS, UAI.
- Reviewing for journals: Journal of Machine Learning Research, Transactions on Machine Learning Research,

Journal of Computational and Graphical Statistics, Bernoulli.

- Co-organizer: Machine Learning and its Applications (2022, Institute for Mathematical Science (NUS) and Simons Institute for the Theory of Computing), Deep learning reading group (2018-2020, Oxford statistics).

Selected Talks

- August 2024: Nvidia, San Francisco (Invited talk).
- May 2024: ThinkAI Hackathon, UM6P, Morocco (Invited talk).
- March 2024: AI Seminar, Uber, San Francisco (Invited talk).
- February 2024: AI Seminar, MBZUAI, Abu Dhabi (Invited talk).
- October 2023: MoDL Seminar, Simons Institute, UC Berkeley (Invited talk).
- June 2023: FLAIR Seminar, EPFL, Switzerland (Invited talk).
- May 2023: Morocco AI Webinar (Invited talk).
- March 2023: Nvidia Research, Hong Kong (Invited talk).
- February 2023: Rising Stars in AI Symposium. KAUST, Saudi Arabia (Invited talk).
- February 2023: AI Seminar at TII, Abu Dhabi (Invited talk).
- January 2023: One World Seminar Series on the Mathematics of Machine Learning (Invited talk).
- December 2022: Computational and Methodological Statistics (CMStatistics 2022), London (Invited talk).
- November 2022: Rough Paths Seminar, The Alan Turing Institute, London (Invited talk).
- September 2022: 3rd Symposium on Machine Learning and Dynamical Systems, Fields Institute, Toronto (Invited talk).
- May 2022: Abu Dhabi Stochastics Seminar, Department of Mathematics, NYU Abu Dhabi (Invited talk).
- Feb 2022: CDSML Seminar Series, National University of Singapore (Invited talk).
- Dec 2021: Seminar Series, Department of Statistics, University of Toronto (Invited talk).
- Dec 2021: Spotlight talk at NeurIPS ICBINB workshop.
- Dec 2020: Department of Mathematics, National University of Singapore (Invited talk).
- Nov 2020: Department of Statistics, University of Oxford (Invited talk).

Teaching

Lecturer

Principles of Machine Learning (DSA5105, graduate level, 100+ enrolled students)

Department of Mathematics, NUS

Semester 1, 2021-2022 & 2022-2023

Advanced Topics in Machine Learning (DSA5202, graduate)

Department of Mathematics, NUS

Semester 2, 2021-2022 & 2022-2023

Class Tutor and Teaching Assistant

Advanced Simulation Methods (SC5, undergraduate)

Department of Statistics, University of Oxford

Hilary term, 2019-2020

Bayesian Inference and MCMC methods (SC7, graduate)

Department of Statistics, University of Oxford

Hilary term, 2018-2019

Other Professional Experience

G-Research

Quantitative Research Intern

London, UK

July 2019 – September 2019

- Trained a neural network model to better estimate the correlation matrix
- Built a neural network model for robust portfolio optimization

Bloomberg LP

Quantitative Research Intern

- Showed (provably) that the empirical correlation matrix overestimates the largest eigenvalue as the dimension of the problem grows (supervised by Bruno Dupire)
- Built a Denoising Autoencoder model to 'clean' the spectrum of the correlation matrix

New York, USA
March 2017 – August 2017

J.P. Morgan

Quantitative Research Intern

- Calibrated a Stochastic model for interest rate products
- Designed an Initial Margin pricer for interest rate swaps

London, UK
April 2016 – September 2016

Kantox

Quantitative Research Intern

- Optimized Margin requirements for FX forwards, a known financial derivative widely traded in the FX market

Barcelona, Spain
June 2015 – September 2015

SNECMA

Part-time Research Intern

- Optimized engine storage facilities using Queuing theory (SNECMA is a leading aircraft engine manufacturer)

Paris, France
September 2014 – April 2015

Technical Skills

Coding

- Python packages: Utils (Numpy, Scipy, Pandas, Seaborn etc.)
- Deep Learning: PyTorch (6+ years), Tensorflow (4+ years), HuggingFace Transformers (2+ years), LangChain (1+ years)
- Cloud Computing: AWS, Google Cloud
- Distributed Training: Accelerate (HuggingFace)

Languages

- Moroccan: native
- Arabic: native
- French: bilingual
- English: fluent
- Spanish: beginner