Sanae Lotfi

New York City, USA
(332) 248 3261
sl8160@nyu.edu
sanaelotfi.github.io

Research Interests

- Theoretical and empirical understanding of generalization in foundation models.
- Probabilistic learning, information theory and model compression.
- Improving reasoning, planning, and problem solving in large language and multimodal models.
- Robustness, reliability, and out-of-distribution generalization.

Education

2020–current **Ph.D. in Data Science, New York University**, USA
- Research focus: Understanding, quantifying, and improving generalization in deep learning.
- Advisor: Andrew Gordon Wilson
- Affiliations: CDS, CILVR

2018–2020 **M.Sc. in Applied Mathematics, Polytechnique Montreal**, Canada
GPA: 4.0/4.0
- Research focus: Stochastic first-, and second-order optimization methods for machine learning.
- Advisors: Andrea Lodi, Dominique Orban
- Affiliations: MILA, CERC, GERAD

2015–2018 **M.Eng. in Applied Mathematics, Centrale Paris**, France
GPA: 3.97/4.33

Publications

2024 Unlocking Tokens as Data Points for Generalization Bounds on Larger Language Models [arxiv]
Sanae Lotfi*, Yilun Kuang*, Brandon Amos, Micah Goldblum, Marc Finzi, Andrew Gordon Wilson
Neural Information Processing Systems (NeurIPS), **spotlight presentation**
ICML Workshop on Theoretical Foundations of Foundation Models, **oral presentation**
Best Paper Award

2024 Non-Vacuous Generalization Bounds for Large Language Models [arxiv]
Sanae Lotfi*, Marc Finzi*, Yilun Kuang*, Tim G. J. Rudner, Micah Goldblum, Andrew Gordon Wilson
International Conference on Machine Learning (ICML)

2023 Bayesian Model Selection, the Marginal Likelihood, and Generalization (Extended Paper) [arxiv]
Sanae Lotfi, Pavel Izmailov, Gregory Benton, Micah Goldblum, Andrew Gordon Wilson
Journal of Machine Learning Research (JMLR), **Best Papers Track**

2023 Mitigating Augmentation Bias with Input-Dependent Distributions over Augmentations
Sanae Lotfi, Tim G. J. Rudner, Brandon Amos, Andrew Gordon Wilson
Under review, soon on arxiv.

2022 PAC-Bayes Compression Bounds So Tight That They Can Explain Generalization [arxiv]
Sanae Lotfi*, Marc Finzi*, Sanyam Kapoor*, Andres Potapczynski*, Micah Goldblum, Andrew Gordon Wilson
Neural Information Processing Systems (NeurIPS)

2022 Bayesian Model Selection, the Marginal Likelihood, and Generalization [pmlr]
Sanae Lotfi, Pavel Izmailov, Gregory Benton, Micah Goldblum, Andrew Gordon Wilson
International Conference on Machine Learning (ICML), **long oral presentation, top 2% submissions**
Outstanding Paper Award

2022 Evaluating Approximate Inference in Bayesian Deep Learning [pmlr]
Andrew Gordon Wilson, **Sanae Lotfi**, Sharad Vikram, Matthew D. Hoffman, Yarin Gal, Yingzhen Li, Melanie F. Pradier, Andrew Foong, Sebastian Farquhar, Pavel Izmailov
NeurIPS Competition and Demonstration Track, Proceedings of Machine Learning Research

Publications Cont.

2022 Adaptive First- and Second-Order Algorithms for Large-Scale Machine Learning [arxiv]
Sanae Lotfi, Tiphaine Bonniot de Ruisselet, Dominique Orban, Andrea Lodi
Annual Conference on Machine Learning, Optimization, and Data Science (LOD), **oral presentation**

2021 Dangers of Bayesian Model Averaging under Covariate Shift [arxiv]
Pavel Izmailov, Patrick Nicholson, **Sanae Lotfi**, Andrew Gordon Wilson
Neural Information Processing Systems (NeurIPS)

2021 Loss Surface Simplexes for Mode Connecting Volumes and Fast Ensembling [arxiv]
Gregory W. Benton, Wesley J. Maddox, **Sanae Lotfi**, Andrew Gordon Wilson
International Conference on Machine Learning (ICML), **spotlight presentation**

2021 Stochastic Damped L-BFGS with Controlled Norm of the Hessian Approximation [arxiv]
Sanae Lotfi, Tiphaine Bonniot de Ruisselet, Dominique Orban, Andrea Lodi
SIAM Conference on Optimization, **oral presentation**

2021 Ocular Cataract Identification Using Deep Convolutional Neural Networks [IEEE]
Feliciana Manuel, Saide Saide, Felermino Ali, **Sanae Lotfi**
International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems

Workshop Papers:

2019 Home Health Care Resource Allocation Problem: A Reinforcement Learning Approach
Sanae Lotfi, Abderrahim Khalifa, Amine Bellahsen, Ola Bdawy, Loubna Benabbou, Ismail El Hallaoui
NeurIPS ML for the Developing World Workshop

2019 Planning in Home Health Care Structures using Reinforcement Learning
Sanae Lotfi, Abderrahim Khalifa, Amine Bellahsen, and Loubna Benabbou
ICLR AI for Social Good Workshop, problem introduction track, **oral presentation**

Thesis:

2020 Stochastic First and Second Order Optimization Methods for Machine Learning
Sanae Lotfi
Master's thesis, Polytechnique Montreal
Best Master's Thesis Award, Department of Mathematics and Industrial Engineering

I served as a research mentor to Feliciana Manuel, whose name is underlined, through the Deep Learning Indaba program. * denotes equal contribution.

Work Experience

June – Sep. 2024 **Research Intern, Microsoft Research**, New York, USA
- Mentors: Jordan T. Ash and Miro Dudík.
- Research on model merging in large language models.

Oct. 2022 – Oct. 2023 **Visiting Researcher, Meta AI, FAIR Labs**, New York, USA
- Mentor: Brandon Amos.
- Research on robustness to model misspecification, learning approximate invariances, improving generalization through input-dependent data augmentations, and generalization bounds for LLMs.
- 3 publications in preparation/under review.

May – Oct. 2022 **Applied Scientist Intern, Amazon AWS**, Santa Clara, USA
- Mentors: Yuyang (Bernie) Wang and Richard Kurle.
- Research on time series modeling under distribution shift.

Feb. – Aug. 2018 **Research Intern, Air Liquide**, Paris, France
- Designing algorithms to predict the gas consumption and optimize the production planning.

July – Dec. 2017 **Research Intern, BeeBryte**, Singapore
- Developing and optimizing strategies for trading and hedging in the electricity markets.

Awards and Honors

2024 **Rising Star in EECS by MIT**
Distinguished as a Rising Star in Electrical Engineering and Computer Science by MIT.

2024 **Best Paper Award at the ICML Theoretical Foundations Workshop**
Awarded for *Unlocking Tokens as Data Points for Generalization Bounds on Larger Language Models* at the ICML Workshop on Theoretical Foundations of Foundation Models.

2023 **Rising Star in Machine Learning Award**
Distinguished as a Rising Star in ML by the University of Maryland Center for Machine Learning.

2023-2025 **Microsoft Research PhD Fellowship**
One of 10 PhD students in Canada and the United States to be awarded this fellowship.

2022 **ICML Outstanding Paper Award**
Awarded to *Bayesian Model Selection, the Marginal Likelihood, and Generalization*.

2022 – 2023 **Meta AI Research Grant**
Covers full tuition and stipend as a part of the Meta AI Mentorship Program.

2022 – 2023 **Travel Awards**
NeurIPS 2023, ICML 2023 (ICML Participation Grant), NeurIPS 2022 (NeurIPS Scholar Award), ICML 2022 (Women in Machine Learning Travel Award).

2021 **Google DeepMind Fellowship**
One of three Google DeepMind Fellows to join NYU in 2020–2021.

2020 – 2025 **Data Science Graduate Fellowship**
5-year graduate fellowship awarded by the NYU Center for Data Science.

2020 **Best Master's Thesis Award**
Awarded by the department of Mathematics and Industrial Engineering at Polytechnique Montréal.

2020 **McKinsey First Generation Achievement Award**
Prize for outstanding individuals who are the first in their family to earn a higher-education degree.

2015 – 2018 **French Government Scholarship for Excellence**
3-year scholarship. Awarded for ranking 2nd in CentraleSupélec's entrance exam.

2013 – 2021 **Académie Hassan II Scholarship for Excellence**
8-year scholarship. Awarded for ranking 1st in nationwide open competition in mathematics.

2010 – 2013 **Various first prizes in regional mathematics and physics Olympiads in Morocco.**

Teaching

Fall '24, '23 and '21 Section Leader, DS-GA 3001: Introduction to Data Science for PhD Students, New York University
- Prepared and delivered lab sessions, created assignments, graded assignments and class projects, and held office hours.

Spring 2022 Grader, DS-GA 1004: Big Data, New York University
- Graded quizzes, assignments and class projects.

Summer 2021 Teaching Assistant, Harnessing Quantum Matter Data Revolution Summer School.
- Prepared and delivered hands-on sessions on density estimation methods, probabilistic PCA, and probabilistic modeling tools.

Fall 2020 Teaching Assistant, IFT6135: Representation Learning, University of Montreal
- Created new quizzes and assignments, graded assignments and exams, and held office hours.

Fall '20, '19, Winter '19 Section Leader, MTH3302: Probability and Statistics for AI, Polytechnique Montréal
- Prepared and delivered lab sessions, created new lab assignments and class projects, graded assignments and exams, and held office hours.

Summer 2018 Mathematics Instructor, Renovo Association
- Developed practice exercises and conducted training sessions to prepare students for rigorous mathematics examinations at the Classes Préparatoires level, targeting highly competitive universities in France.

2016 - 2018 Mathematics and Physics Instructor, Renovo Association
- Created and delivered mathematics and physics to students from disadvantaged backgrounds at the high school level.

Invited Talks

Are the Marginal Likelihood and PAC-Bayes Bounds the Right Proxies for Generalization?

Nov. 2023 Rising Stars in Machine Learning Workshop, UMD
Oct. 2023 Carnegie Mellon University (CMU), Artificial Intelligence Seminar Series
Mar. 2023 Massachusetts Institute of Technology (MIT), CSAIL Seminar
Mar. 2023 Harvard University, Data to Actionable Knowledge Lab
Feb. 2023 FAIR Labs, Meta AI NYC
Dec. 2022 North Africans in ML Workshop at NeurIPS, **Keynote Speaker**

Non-Vacuous Generalization Bounds for Large Language Models

Dec. 2024 NeurIPS Workshop on Compression and Information Theory, **Keynote Speaker** (upcoming)
June 2024 ML Collective, Deep Learning: Classics and Trends
May 2024 ML Reading group at the University of Illinois
Mar. 2024 Cohere For AI, Guest Talk

PAC-Bayes Compression Bounds So Tight That They Can Explain Generalization

June 2023 FAIR Labs, Meta AI NYC
Mar. 2023 New York University, CDS Graduate Student Seminar
Dec. 2022 Neural Information Processing Systems (NeurIPS)
Nov. 2022 Women in AI Ignite at NeurIPS, Invited Talk

Bayesian Model Selection, the Marginal Likelihood, and Generalization

May 2023 The Data Science and Machine Learning Research Group
Sep. 2022 ML Collective, Deep Learning: Classics and Trends
July 2022 Amazon, Forecast Science Talks
July 2022 International Conference on Machine Learning (ICML), **Long Oral**, [video]
June 2022 INRIA Social Data Group
Apr. 2022 Morocco AI, Webinar Series [video]
Apr. 2022 New York University, CDS Graduate Student Seminar

Mitigating Augmentation Bias with Input-Dependent Distributions over Augmentations

July 2023 Generative AI Lightning Talk, FAIR Labs, Meta AI

Robustness of Deep Learning Models to Distribution Shift

July 2022 Women in Machine Learning Workshop at ICML, **Session Leader**

Understanding and Quantifying Generalization in Deep Learning Models

Oct. 2021 Women in Data Science Panel, New York University
Sep. 2021 DeepMind Montreal

Dangers of Bayesian Model Averaging under Covariate Shift

Dec. 2021 Neural Information Processing Systems (NeurIPS), [video]
Nov. 2021 Women in Mathematics Research Talks, New York University

Adaptive First and Second Order Algorithms for Large-Scale Machine Learning

July 2021 SIAM Conference on Optimization
Dec. 2020 NeurIPS Optimization for ML Workshop, **Spotlight Presentation**, [video]
Nov. 2019 Montreal Machine Learning and Optimization Group

Planning in Home Health Care Structures using Reinforcement Learning

May 2019 ICLR AI for Social Good, **Spotlight Presentation**

Invited Panels

Mar. 2022 Affinity Group Supported Pathways to ML Research Panel and Social, International Conference on Artificial Intelligence and Statistics (AISTATS)

Oct. 2021 Data Science Career Panel, Women in Data Science (WiDS) at NYU

Professional and Community Activities

Organizing and Leadership

2024 Organizer of the NeurIPS Workshop on *Scientific Methods for Understanding Neural Networks*

2023 Organizer of the NeurIPS Muslims in ML Affinity Workshop

2022 Leader of the "Robustness of Deep Learning Models to Distribution Shift" session at the Women in Machine Learning Workshop, ICML

2021 Organizer of the NeurIPS Competition on *Approximate Inference in Bayesian Deep Learning*

2021 Founding organizer of research talks, *Tea-Talks*, at the NYU Association of Women in Mathematics

2016 Co-founder of the competitive programming association at Centrale Paris

Outreach and Volunteering

2022 Mentor for the NeurIPS High School Outreach Program to broaden participation in machine learning by engaging with students early

2022 Volunteer at the Women in Machine Learning Workshop, NeurIPS

2021-2022 Mentor for the Deep Learning Indaba Mentorship Programme to support and strengthen the African machine learning community

2020 Consultant for the university of Montreal to increase enrollment in mathematics by students from underrepresented groups

2016-2018 Mathematics and physics volunteer instructor for high school students from disadvantaged backgrounds

Research Mentorship

2021 – 2022 Mentored Feliciana Manuel to conduct her undergraduate research on "Ocular Cataract Identification Using Deep Convolutional Neural Networks" at Lúrio University, Mozambique.

- Paper accepted to the International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems.

Reviewing

Conferences Neural Information Processing Systems (NeurIPS), International Conference on Machine Learning (ICML), International Conference on Learning Representations (ICLR), Conference on Machine Learning, Optimization, and Data Science (LOD)

Journals Journal of Machine Learning Research (JMLR)

Workshops NeurIPS 2023 Attributing Model Behavior at Scale Workshop, NeurIPS 2022 Women in Machine Learning Workshop - Area Chair, NeurIPS 2021 Bayesian Deep Learning Workshop, NeurIPS 2019 Women in Machine Learning Workshop

Competitions NeurIPS 2023 Competition Track

Mental Health Advocacy

2023 Organizer of *CDS Peers Meet-ups*, where students from the Data Science department at NYU meet to support and mentor each other in navigating the challenges that arise during the PhD.

2023 Co-organizer and coordinator of the workshop "Everything I wish I knew as I navigated my Ph.D." at NYU and CMU, where we discussed mental health challenges, academic culture in doctoral programs, and how to create mental health advocacy groups and activities in both universities.

2017 Mental health campaign leader: led a group of over 70 students to promote emotional well-being and prevent suicide among hundreds of students at Ecole Centrale Paris.

Technical Skills

Proficient	Python (Scikit-learn, SciPy stack, PyTorch), Git, Latex
Experienced	TensorFlow, Julia, Matlab, R

Technical Reports

2022 Understanding the Generalization of Deep Neural Networks through PAC-Bayes bounds [report]
Joint with Andres Potapczynski, Anthony Chen, and Chris Ick

2021 Causal Representation Learning [report]
Joint with Taro Makino and Lily Zhang

2019 Analysis of High Dimensional Distributions with Decoupled Norm and Direction [report]
Joint with Jose Gallego, Ankit Vani, and Max Schwarzer

2019 Variance Reduction with Neighbours for Adaptive Optimization [report]
Joint with Jose Gallego

Summer Schools

2024 Chatting Minds: The Science and Stakes of Large Language Models, Hybrid Summer School

2023 Oxford Machine Learning Summer School: Machine Learning Fundamentals, Cases, and Health Care Applications

2021 Deep Learning Theory Summer School at Princeton

2021 Harnessing Quantum Matter Data Revolution, Virtual Summer School

Selected Media Coverage

2024 *Do Large Language Models Really Generalize? This Paper Says Yes*, NYU Center for Data Science

2021 *Scholar Q&A: Sanae*, DeepMind

2020 *DeepMind Fellow Profile: Sanae Lotfi*, NYU Center for Data Science

2016 *Barcelonnette vise l'autonomie énergétique*, Magazine Barcelonnette