

Research Interests

Large Language Models, Diffusion Models, Self-Supervised Learning, Vision-Language Models, Probabilistic Generative Models, AI for Science, Generalization Theory, Numerical Methods

Education

2023–present **Ph.D. in Data Science, New York University, USA.**

- Research Focus: Large Language Models, Diffusion Models, AI for Science, SSL
- Advisor: [Andrew Gordon Wilson](#)
- Affiliation: [CDS](#), [CILVR](#)

2020–2023 : **B.A. in Mathematics with High Honors, Courant Institute, New York University, USA.**

- Latin Honors: Magna Cum Laude with Minor in Computer Science; 8 PhD-Level Courses in ML

Publications

Conference Papers

- 2025 **Customizing the Inductive Biases of Softmax Attention using Structured Matrices.**
Yilun Kuang, Noah Amsel, Sanae Lotfi, Shikai Qiu, Andres Potapczynski, Andrew Gordon Wilson.
Under Review
- 2024 **Bayesian Optimization of Antibodies Informed by a Generative Model of Evolving Sequences.**
Alan Nawzad Amin, Nate Gruver*, **Yilun Kuang*** (equal contribution), Yucen Lily Li*, Hunter Elliott, Aniruddh Raghu, Calvin McCarter, Peyton Greenside, Andrew Gordon Wilson.
International Conference on Learning Representations (ICLR), 2025, **Spotlight**
- 2024 **Unlocking Tokens as Data Points for Generalization Bounds on Larger Language Models.**
Sanae Lotfi*, **Yilun Kuang*** (equal contribution), Brandon Amos, Micah Goldblum, Marc Finzi, Andrew Gordon Wilson.
Neural Information Processing Systems (NeurIPS), 2024, **Spotlight**
- 2023 **Non-Vacuous Generalization Bounds for Large Language Models.**
Sanae Lotfi*, Marc Finzi*, **Yilun Kuang*** (equal contribution), Tim G. J. Rudner, Micah Goldblum, Andrew Gordon Wilson.
International Conference on Machine Learning (ICML), 2024
- 2023 **Learning Efficient Coding of Natural Images with Maximum Manifold Capacity Representations.**
Thomas Yerxa, **Yilun Kuang**, Eero Simoncelli, SueYeon Chung.
Neural Information Processing Systems (NeurIPS), 2023

Workshop Papers

- 2024 **Bayesian Optimization of Antibodies Informed by a Generative Model of Evolving Sequences.**
Alan Nawzad Amin, Nate Gruver*, **Yilun Kuang*** (equal contribution), Yucen Lily Li*, Hunter Elliott, Aniruddh Raghu, Calvin McCarter, Peyton Greenside, Andrew Gordon Wilson.
NeurIPS 2024 Workshop: AI for New Drug Modalities, **Spotlight**
- 2024 **Unlocking Tokens as Data Points for Generalization Bounds on Larger Language Models.**
Sanae Lotfi*, **Yilun Kuang*** (equal contribution), Brandon Amos, Micah Goldblum, Marc Finzi, Andrew Gordon Wilson.
ICML 2024 Workshop: Theoretical Foundations of Foundation Models, **Best Paper Award**

- 2023 **Unsupervised Learning on Spontaneous Retinal Activity Leads to Efficient Neural Representation Geometry.**
 Andrew Ligeralde*, **Yilun Kuang*** (equal contribution), Thomas Edward Yerxa, Miah N Pitcher, Marla Feller, SueYeon Chung.
 NeurIPS 2023 Workshop: Unifying Representations in Neural Models (UniReps)
- 2023 **Non-Vacuous Generalization Bounds for Large Language Models.**
 Sanae Lotfi*, Marc Finzi*, **Yilun Kuang*** (equal contribution), Tim G. J. Rudner, Micah Goldblum, Andrew Gordon Wilson.
 NeurIPS 2023 Workshop: Self-Supervised Learning & Mathematics of Modern Machine Learning (M3L)
[Poster](#)
- 2023 **Learning a Visual Representation by Maximizing Manifold Capacity.**
 Thomas Yerxa, **Yilun Kuang**, Eero Simoncelli, SueYeon Chung.
 Computational and Systems Neuroscience (COSYNE), 2023

Work Experience

- June – Sep 2022 **Research Intern at Flatiron Institute, Simons Foundation.**
 o Mentor: [SueYeon Chung](#)
 o Develop a state-of-the-art self-supervised learning algorithm for vision.
 o 1 publication in NeurIPS and 1 poster presentation in COSYNE.
- Jan 2023 **Open Source Contributor at MosaicML.**
 o Implement a distributed training pipeline for efficient deep neural network training. [Open source contribution](#) available in the Composer Library documentation.

Fellowships & Awards

- 2023 – 2028 **Center for Data Science Fellowship.**
 o Awarded the NYU Center for Data Science Fellowship for five years and the Data Science Supplementary Fellowship Grant
- May 2023 **Nicholas and Andrea Ferrara Research Scholar, Dean's List (2020-2022), DURF Grant Recipient, Best Presenter at NYU Undergraduate Research Conference.**
- May 2023 **Meritorious Winner in 2021 Mathematical Contest in Modeling (MCM) (Top 7%).**

Teaching Assistantship

- Fall 2024 **Grader**, CSCI-GA 2565: Machine Learning, New York University.
- Spring 2024 **Grader**, DS-GA 1003: Machine Learning, New York University.
- Fall 2023 **Section Leader & Grader**, DS-GA 1011: Natural Language Processing with Representation Learning, New York University.
- Spring 2022 **Grader**, DS-GA 1012: Natural Language Understanding and Computational Semantics, New York University.

Summer School

- Summer 2020 **Neuromatch Academy in Computational Neuroscience**, Virtual.
- Summer 2020 **Summer Session in Mathematics**, Harvey Mudd College, CA.

Project Report

- Dec 2022 **A Survey of Double Descent in High-Dimensional Linear Regression**, [\[Link\]](#).
- May 2022 **A Survey of Lazy and Feature Learning Regimes**, [\[Link\]](#).
- May 2021 **Scale-Invariant Finetuning for Improved Generalization**, [\[Link\]](#).

Position of Responsibility

Conference **Reviewer**, ICML 2025, NeurIPS 2024, ICLR 2025, AISTATS 2025.

Workshop **Reviewer**, NeurIPS 2023 Workshop: Self-Supervised Learning - Theory and Practice, ICML 2024 Workshop: Theoretical Foundations of Foundation Models (TF2M), NeurIPS 2024 Workshop: Adaptive Foundation Models: Evolving AI for Personalized and Efficient Learning (AFM).

Media Coverage

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

Computer skills

Programming Languages Python, C++, C, Java, Matlab, Julia, SQL, Bash, Vim, LaTeX

Machine Learning PyTorch, Jax, Hugging Face, TensorFlow, Scikit-Learn, GPyTorch, NumPy, Pandas

Cloud Services/Other AWS, GCP, Docker, Kubernetes, Vim, Git, MicroService, CUDA, OpenMP, MPI