

Yukai Yang

[Website](#) | [Email](#) | [LinkedIn](#) | [GitHub](#)

EDUCATION

University of Chicago

M.S. in Statistics

Chicago, IL

Sep 2023 - Mar 2025 (expected)

- GPA: 3.7
- Research Interest: Uncertainty Quantification, High-dimensional Statistics, Machine Learning Theory

New York University

B.A. in Data Science and Mathematics

New York, NY

Aug 2019 - Dec 2022

- Graduated with Magna Cum Laude
- Courses taken: Foundations of Machine Learning, Deep Learning, Mathematical Statistics, ML for Functional Genomics (Columbia), Advanced Techniques in DL, Convex and Non-smooth Optimization, etc.

PEER-REVIEWED PAPERS

* indicates equal contributions.

Workshop Papers

- [1] Yukai Yang*, Tracy Zhu*, Marco Morucci, Tim G. J. Rudner. **Weak-to-strong Confidence Prediction.** Workshop on Statistical Foundations of Large Language Models, Attributing Model Behavior at Scale, Safe Generative AI, and Regulatable ML. (**NeurIPS Workshop**), 2024

RESEARCH & TEACHING EXPERIENCE

Visiting Researcher (Remote)

Department of Statistics and Data Science, Yale University

Sep 2024 -

New Haven, CT

- Analyze representation learning behavior of attention-based models with low-rank adaptation
- Provide theoretical understanding of how learning rate and chosen rank impact the model's generalization

Student Researcher (Remote)

Center for Data Science, New York University

Feb 2024 -

New York, NY

- Study hallucinating behaviors of large language models from their answer uncertainty
- Train weaker models to generalize and to predict strong models' performance (weak-to-strong generalization)

Visiting Researcher (Remote)

Trustworthy AI Lab, University of California

Mar 2024 -

Los Angeles, CA

- Align Large Language Models with safety standards using in-context learning
- Train a prompt attacker/defender system in an adversarial manner to optimize the prompts for safety

Visiting Academics: Research Assistant

Center for Data Science, New York University

Feb 2022 - Oct 2023

New York, NY

- Use deep bayesian active learning to select the most informative images from a unlabeled pool
- Integrate function-space regularization with acquisition functions to improve uncertainty quantification

Teaching Assistant and Graders

2022, 2023, 2024

- University of Chicago: Clinical Data Science (DS, TA)
- New York University: Linear Algebra (Math, TA), Advanced Topics in Data Science: Deep Learning (DS, Grader)

PROJECTS

HAAT: Improve Adversarial Robustness via Hessian Approximation | *Adversarial Training*

- Improve Projected Gradient Descent with our new algorithm that includes a second-order approximation term
- Analyze and explain different PGD-based algorithms improvements with learning theory techniques

Uniform Convergence for Double-descent Curve in Different Models | *Statistics, DL Theory*

- Give a survey-style summary of recent relevant work on theoretical analysis of double-descent curve
- Provide a proof for the bound of convergence and experiment with random feature models and DNNs

Benchmarking GPT-3: how LLMs Learn New Words | *NLP, Few-shot Learning with LLMs*

- Examine whether LLMs manage to reason in a similar way as an intellectual agent to understand new words
- Help models to learn complicated compound words faster | Prevent inadvertent entry from confusing the model

Using Function-Space-VI-based Active Learning to Label Protest Images | *CV, Active Learning*

- Apply and improve function space variational inference as the heuristic functions adaptable to the dataset
- Develop pretrained AL model to help annotate protest images more efficiently

Kernel Approximation for Gradient Descent based Algorithms | *Convex Optimization*

- Explore how kernel methods can be used as an equivalence or approximation of various gradient descent algorithms
- Improve the error bounds of the approximations to better interpret how deep learning models learn features

GRANTS

Summer Research Grant - \$5000

Jul 2024

Center for Data Science, New York University

New York, NY

- Award kindly provided by Tim G. J. Rudner
- Work on weak-to-strong prediction of Large Language Models' uncertainty

Wasserman Center Internship Grant - \$1000

Jun 2022

Wasserman Center for Career Development, New York University

New York, NY

- Summer internship for students to conduct research
- Studied active learning with entropy-based heuristic for vision models

COMPETITIONS

ICPC in Greater New York Region

Mar 2022

International Collegiate Programming Contest

New York, NY

- Programming Competition that relies on programming and math
- 1st place in all NYU team. 3rd place in the Greater New York Region
- Heading to the North American Competition

Bud Challenge 2021 – Data Science track

Jun 2021

Business competition held by the beer company Budweiser

Shanghai

- Build a B2B, ML-based recommendation system for retailers.
- 3rd Place in the Final Round.

TECHNICAL SKILLS

Programming Languages: Python, R, SQL, Bash, MATLAB, C.

Libraries: PyTorch, TensorFlow, Jax, transformers, opencv, H2O, ray, nltk, scipy, pandas, matplotlib, seaborn, numpy

Other Tools: GCP, HPC, AWS, Postgres, MongoDB