

Xuelin Yang

xuelin@berkeley.edu <https://xyang23.github.io/>

EDUCATION

University of California, Berkeley

Ph.D. Student, Statistics

Berkeley, CA

08/2023 – Present

Stanford University

Stanford, CA

Master of Science, Computer Science (AI track), GPA: 4.0

11/2021 – 06/2023

- **Relevant Coursework:** Graphics, Natural Language Processing, Computer Vision, Machine Learning.

Bachelor of Science, Mathematics, GPA: 3.99

09/2019 – 06/2023

- **Relevant Coursework:** Algebra, Analysis, Stochastic Process, Optimization.
- **8 A+ courses** in Algebraic Number Theory, Complex Analysis, Applied Matrix Theory, Probability Theory, Stochastic Process, Computer Graphics, and Natural Language Understanding.

RESEARCH INTEREST

My research interest lies broadly in **Applied Statistics and Machine Learning**. One of my current aspirations is to unravel the complexity of market interactions and behaviors in **Econ ML**. Specifically, I am excited to 1) explore how we understand the market from a macro and micro perspective through the lens of statistical models, and 2) uncover hidden patterns and correlations that might provide invaluable guidance for policymakers, investors, and stakeholders navigating the intricate landscape of markets.

EXPERIENCE

Research Assistant, Stanford Statistics (advised by Robert Tibshirani) Stanford, CA, 01/2022 – Present

- Propose FastCPH, the first linear-time hazard function for survival prediction on health data.
- Advance the SOTA performance in multiple survival datasets by FastCPH networks with feature sparsity.
- Introduce boosting model co-training for cooperative learning on multi-view biomedical analysis.

Research Assistant, MIT CSAIL (advised by Yoon Kim)

Cambridge, MA, 06/2022 – 03/2023

- Develop probabilistic metrics on sampling surrounding distribution and Kolmogorov complexity and achieve promising zero-shot evaluations of semantic similarity.
- Introduce novel approaches to analyze language models in functional space with Neural Tangent Kernel (NTK) as the first NLP application of the renowned NTK theorem.

Research Assistant, Stanford AI Lab (advised by Jiajun Wu)

Stanford, CA, 06/2021 – 06/2022

- Construct CLEVRER-Humans, a novel subjective causal relation video question-answering dataset.
- Introduce iterative cloze-based annotation of event descriptions and hybrid event description generation using neural sequence-to-sequence models.
- Propose CLEVRER-Humans as a unique and important challenge of combined physical scene understanding, natural language understanding, and causal reasoning by zero-shot/few-shot evaluations.

Research Assistant, Stanford Info Lab (advised by Jure Leskovec)

Stanford, CA, 09/2020 – 06/2021

- Introduce Bayesian inference for neuro-symbolic zero-shot concept recognition and acquisition.
- Construct an energy-based learning paradigm for visual concept learning and zero-shot generalization.
- Develop novel models for hierarchical concepts detection given graph-based symbolic representations.

Software Engineering Intern, TikTok

Mountain View, CA, 06/2020 – 09/2020

- Build the image-to-comic TikTok special effect that achieved 2 billion downloads with the VR group.
- Significantly improve image-to-image translation with StyleGAN and vector quantization.

PUBLICATIONS / PATENTS

Mao, J.*, Yang, X.*, Zhang, X., Goodman, N., Wu, J. (2022) CLEVRER-Humans: Describing Physical and Causal Events the Human Way. *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS 2022 Dataset Track)*. *: Contributed equally. Alphabetical order.

Yang, X., Abraham, L., Kim, S., Smirnov, P., Ruan, F., Haibe-Kains, B., Tibshirani, R.. (2022) FastCPH: Efficient Survival Analysis for Neural Networks. *Workshop on Learning from Time Series for Health, Workshop on Algorithmic Efficiency in Practical Neural Network Training, 36th Conference on Neural Information Processing Systems (2 NeurIPS 2022 Workshop Posters)*.

Wu, T., Tjandrasuwita, M., Wu, Z., **Yang, X.,** Liu, K., Sosić, R., & Leskovec, J. (2022). Zeroc: A neuro-symbolic model for zero-shot concept recognition and acquisition at inference time. *Advances in neural information processing systems (NeurIPS 2022)*. (also appeared as **ICML 2022 Workshop Poster** for *Beyond Bayes: Paths Towards Universal Reasoning Systems Workshop*).

DISTINCTIONS

Candidate, **Tau Beta Pi (TBP) Engineering Honor Society**, 2023 (top fifth of all seniors in the School of Engineering at Stanford based on GPA)

Second place at Stanford, **Jane Street Estimathon**, 2020 (top 2 out of 20)

3rd Place Award in System Software, **Intel International Science and Engineering Fair**, 2018 (top 10% out of 1800 finalists, state's best record in 8 years, the world's largest pre-college STEM competition)

First Prize, **Provincial Adolescent Science and Technology Innovation Contest**, 2018 (top 5 out of 400 in the province)

Best High School Student of the Year, awarded by All-China Students' Federation, 2017 (top 10 out of 170k in the country)

Bronze Award in World's Final, **Shing-Tung Yau Science Award**, 2017 (top 0.5% out of 5800, Chinese largest STEM competition)

Meritorious, **International Mathematical Modelling Challenge**, 2016 (top 10 out of 300 in the country)

EXTRACURRICULAR ACTIVITIES

President, **Association of Chinese Students and Scholars at Stanford**, 2021-22

Member, **Stanford Dragon Boat Club**, 2019-21