Yilan Chen

University of California San Diego, La Jolla, CA

yic031@ucsd.edu · (+1) 858-260-9293 · yilanchen6.github.io

Education

University of California San Diego (UCSD) — San Diego, USA
Ph.D. in Computer Science — *Sep. 2022 - Dec. 2026 (expected)*

Advisors: Prof. Arya Mazumdar & Prof. Yian Ma
Committee members: Mikhail Belkin, Sanjoy Dasgupta, Sicun Gao

University of California San Diego (UCSD) — San Diego, USA
M.S. in Computer Science — *Sep. 2020 - Jun. 2022*

Advisors: Prof. Tsui-Wei (Lily) Weng & Dr. Lam M. Nguyen (IBM Research)

Xi'an Jiaotong University (XJTU) — Xi'an, China
B.E. in Information Engineering — *Aug. 2015 - Jun. 2019*

Research Interests

Areas of Interest: Machine Learning, Deep Learning, Large Language Models, Foundation Models
Tools: Linear Algebra/Matrix Analysis, High Dimensional Statistics, (Non-) Convex Optimization

Publications

A Unified Framework of Post-training Language Models with Reinforcement Learning and Distillation.

- **Yilan Chen**, Ankit Singh Rawat, Aditya Krishna Menon.
- In submission.

Generalization Bound of Gradient Flow through Training Trajectory and Data-dependent Kernel. [link]

- **Yilan Chen**, Zhichao Wang, Wei Huang, Andi Han, Taiji Suzuki, Arya Mazumdar.
- Thirty-Ninth Annual Conference on Neural Information Processing Systems (NeurIPS 2025).

How Does Label Noise Gradient Descent Improve Generalization in the Low SNR Regime? [link]

- Wei Huang[1], Andi Han[1], Yujin Song, **Yilan Chen**, Denny Wu, Difan Zou, Taiji Suzuki.
- Thirty-Ninth Annual Conference on Neural Information Processing Systems (NeurIPS 2025).

Provable and Efficient Dataset Distillation for Kernel Ridge Regression. [link]

- **Yilan Chen**, Wei Huang, Tsui-Wei Weng.
- Thirty-eighth Conference on Neural Information Processing Systems (NeurIPS 2024).

Cross-Task Linearity Emerges in the Pretraining-Finetuning Paradigm. [link]

- Zhanpeng Zhou[1], Zijun Chen[1], **Yilan Chen**, Bo Zhang, Junchi Yan.
- Forty-first International Conference on Machine Learning (ICML 2024).

Analyzing Generalization of Neural Networks through Loss Path Kernels. [link]

- **Yilan Chen**, Wei Huang, Hao Wang, Charlotte Loh, Akash Srivastava, Lam M. Nguyen, Tsui-Wei Weng.
- Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS 2023).

Analyzing Deep PAC-Bayesian Learning with Neural Tangent Kernel: Convergence, Analytic Generalization Bound, and Efficient Hyperparameter Selection. [link]

- Wei Huang[1], Chunrui Liu[1], **Yilan Chen**, Richard Yi Da Xu, Miao Zhang, Tsui-Wei Weng.
- Transactions on Machine Learning Research (TMLR 2023).

The Importance of Prompt Tuning for Automated Neuron Explanations. [link]

- Justin Lee[1], Tuomas Oikarinen[1], Arjun Chatha, Keng-Chi Chang, **Yilan Chen**, Tsui-Wei Weng.
- NeurIPS 2023 Workshop on Attributing Model Behavior at Scale.

Quantifying the Knowledge in a DNN to Explain Knowledge Distillation for Classification. [link]

- Quanshi Zhang[1], Xu Cheng[1], **Yilan Chen**, Zhefan Rao.
- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI 2023).

On the Equivalence between Neural Network and Support Vector Machine. [link]

- **Yilan Chen**, Wei Huang, Lam M. Nguyen, Tsui-Wei Weng.
- Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS 2021).

Explaining Knowledge Distillation by Quantifying the Knowledge. [link]

- Xu Cheng, Zhefan Rao[2], **Yilan Chen**[2], Quanshi Zhang.
- 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020).

Experiences

Google Research New York, USA

Improving the Post-training of Large Language Models with Knowledge Distillation *Jun. 2025 - Sep. 2025*

Student Researcher [Hosts: Ankit Singh Rawat, Aditya Krishna Menon]

- **Pretrained** Gemma-3 models on Wikipedia and SlimPajama datasets with knowledge distillation from larger models;
- **Supervised-finetuned** Gemma-3-1b and Gemma-3-4b on Tulu3 datasets with **knowledge distillation**;
- Trained Gemma-3-1b, Qwen3-1.7B, Qwen2.5-1.5B-Instruct with **reinforcement learning** algorithms such as **GRPO** to imporve the models' reasoning ability. Implemented recent **verifier-free** algorithm, which performs better than GRPO;
- Proposed a **unified framework** to improve GRPO with **knowledge distillation**, which achieves 9.4% higher accuracy than GRPO and 4.9% higher than the KDRL baseline on AIME 2025.

University of California San Diego San Diego, USA

Deep Learning Theory and Its Applications, Trustworthy Machine Learning *Jan. 2021 - present*

Graduate Student Researcher [Advisors: Prof. Arya Mazumdar, Yian Ma, Lily Weng]

- Established equivalence between wide neural networks and support vector machines with **Neural Tangent Kernel** (NeurIPS 2021);
- Established equivalence between general neural networks and kernel machines with a proposed new kernel called **loss path kernel**;
- Derived tight **generalization bounds** of neural networks using the above equivalence. Applied the results to guide the design of **neural architecture search (NAS)** and demonstrate favorable performance compared with SOTA NAS algorithms (NeurIPS 2023);
- For **dataset distillation** of kernel ridge regression, proved theoretically that one data per class is necessary and sufficient to recover the original model's performance in many settings. Proposed **provable and efficient algorithm** for dataset distillation (NeurIPS 2024).

The University of Tokyo and RIKEN AIP Tokyo, Japan

Theory for Large Language Models and Deep Learning *Jun. 2024 - Sep. 2024*

Research Intern [Advisor: Prof. Taiji Suzuki]

- Developed theoretical framework for emerging abilities of LLMs including **in-context learning, scaling law, and chain-of-thought**;
- Derived sharp generalization bounds of neural networks through a connection with kernel method and using **stability** property of gradient descent. The proved bound matches the bound of classic kernel methods (NeurIPS 2025).

Shanghai Jiao Tong University Shanghai, China

Interpretable Machine Learning *Jul. 2019 - Jun. 2020*

Research Intern [Advisor: Prof. Quanshi Zhang]

- Proposed a method to interpret the success of **knowledge distillation** by quantifying and analyzing the task-relevant and task-irrelevant **visual concepts** that were encoded in intermediate layers of a deep neural network (DNN);
- Developed three hypotheses explaining why knowledge distillation performs better than standard training. Created three types of mathematical metrics to assess the feature representations of the DNN and validate the hypotheses.
- Performed extensive experiments to diagnose various DNNs and verified all three hypotheses (CVPR 2020 & TPAMI 2023).

Invited Talks

Generalization Bound of Gradient Flow through Training Trajectory and Data-dependent Kernel

- EnCORE Institute Workshop on Theoretical Perspectives on LLMs *Mar 2025*

Analyzing Neural Networks through Equivalent Kernels

- RIKEN AIP – SJTU CS Joint Workshop on Machine Learning and Brain-like Intelligence *Aug 2024*
- Southern California Applied Mathematics Symposium (SOCAMS 2024) *Apr. 2024*

Analyzing Generalization of Neural Networks through Loss Path Kernels

- ByteDance *Jan. 2024*
- AI TIME *Nov. 2023*

Teaching

DSC 212: Probability and Statistics for Data Science, Teaching Assistant *Fall 2024*
DSC 140B: Representation Learning, Teaching Assistant *Spring 2024*
DSC 210: Numerical Linear Algebra, Teaching Assistant *Fall 2023*
DSC 291: Trustworthy Machine Learning, Teaching Assistant *Fall 2021*

Professional Service

Conference Reviewer: ICML (2022, 2023, 2024, 2025), NeurIPS (2022), ICLR (2022, 2024, 2025)
Journal Reviewer: Journal of Optimization Theory and Applications (JOTA)

Technical Skills

Languages: Python, C/C++, MATLAB, JavaScript
Machine Learning: PyTorch, JAX
Miscellaneous: Linux, LaTex, FPGA, ARM

Honors & Awards

NeurIPS 2023 Scholar Award *Dec. 2023*
Simons Institute Deep Learning Theory Workshop Funding *Aug. 2022*
Outstanding Student Award *Sep. 2016 and 2018*
"Siyua" Merit Scholarship *Sep. 2016, 2017, and 2018*