Shirui (Carl) Chen

Email LinkedIn GitHub Google Scholar
Seattle, WA 2068228936

Applied Mathematics Ph.D. Candidate with 5+ years of experience in machine learning, generative models, and large-scale deep learning. Proven impact in scaling models to billions of parameters, reducing error rates, and building efficient training/inference pipelines. Strong PyTorch/Python practitioner with statistical modeling and signal processing background.

RESEARCH INTERESTS

Generative Models, Neural Network Generalization, Reinforcement Learning, Computational Neuroscience, Machine Learning.

EDUCATION

University of Washington, Seattle 2021 - 2026 (expected)
Ph.D. Candidate Advisor: Prof. Lillian Ratliff (ECE), Prof. Eric Shea-Brown (AMATH)
M.S., Applied Mathematics Major GPA: 4.0

University of Wisconsin, Madison 2017 - 2021
B.Sc. in Mathematics and Computer Science Major GPA: 4.0
Nominated for 2021 Dean's Prize (top 51 students in the class of 2021)

INDUSTRY EXPERIENCE

Research Scientist Intern 2024 June - 2024 Dec
Meta *New York, NY*

- Predicted user gestures and handwriting characters from electromyography (EMG) signals using convolutional and transformer networks.
- Used an encoder–decoder architecture to enable online classification.
- Scaled the model to 1B parameters with grouped query attention (GQA), RMS normalization, and mixed-precision training, reducing the character error rate in handwriting prediction by 39%.
- Employed a Gumbel–softmax layer to learn discretized EMG signal representations, reducing the false positive rate on null behaviors by nearly 100%.
- Implemented a global-local attention mask and torch flex attention to achieve a 40× increase in the size of context window.

Enterprise Software Engineer Intern 2019 May - 2019 August
Facebook Inc. *Menlo Park, CA*

- Developed an intuitive web interface enabling users to seamlessly create and modify purchase orders. Ensured real-time synchronization of changes with the backend database, leading to the project's widespread use within the organization. The implemented interface greatly simplified the enterprise ordering pipeline.
- Full-stack role involves React, JavaScript, Relay, and Hack (PHP).
- Contributed over 3000 lines of code and received a return offer for superior performance.

PUBLICATION

Peer-reviewed work

[1] **Shirui Chen**, Linxing Jiang, P.N. Rajesh Rao, and Eric Shea-Brown. "Expressive probabilistic sampling in recurrent neural networks". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2023. URL: https://arxiv.org/abs/2308.11809.

[2] **Shirui Chen**, Stefano Recanatesi, and Eric Shea-Brown. "A simple connection from loss flatness to compressed representations in neural networks". In: *ICML 2025, 3rd Workshop on High-dimensional Learning Dynamics (HiLD)* (2025). URL: https://arxiv.org/abs/2310.01770.

[3] **Shirui Chen**, Qixin Yang, and Sukbin Lim. "Efficient inference of synaptic plasticity rule with Gaussian process regression". In: *iScience* 26.3 (2023), p. 106182. ISSN: 2589-0042. DOI: 10.1016/j.isci.2023.106182.

[4] Linxing Preston Jiang, **Shirui Chen**, Emmanuel Tanumihardja, Xiaochuang Han, Weijia Shi, Eric Shea-Brown, and Rajesh PN Rao. "Data Heterogeneity Limits the Scaling Effect of Pretraining Neural Data Transformers". In: *COLM 2025, LM4Sci Workshop* (2025), pp. 2025–05.

[5] **Shirui Chen**, Kai Zhou, Liguang Yang, Guohui Ding, and Hong Li. "Racial Differences in Esophageal Squamous Cell Carcinoma: Incidence and Molecular Features". In: *BioMed Research International* 2017 (Mar. 14, 2017). Publisher: Hindawi, p. 1204082. ISSN: 2314-6133. DOI: 10.1155/2017/1204082.

Preprint

[6] David Hsu, Mohsen Mazrooyisebdani, Lucas Alan Sears, Anshika Singh, Mateo N Silver, Makayla Boyarski, Murielle Hsu, **Shirui Chen**, and Raheel Ahmed. "Revisiting linear regression of dynamical systems within the context of Zwanzig-Mori theory: tests on a simple system". In: (Oct. 2021). DOI: 10.36227/techrxiv.16864543.v1.

Abstract/Poster

[7] **Shirui Chen**, Qixin Yang, and Sukbin Lim. "Efficient inference of synaptic plasticity rule with Gaussian process regression". In: *COSYNE abstract* (2022).

[8] **Shirui Chen**, Preston Jiang, Rajesh Rao, and Eric Shea-Brown. "Expressive probabilistic sampling in recurrent neural networks". In: *9th Annual BRAIN Initiative Meeting* (2023).

SELECTED RESEARCH PROJECTS

Large-Scale Pretraining of the π_0 Model on the SO-100 Robot Platform

- Engineered a data curation pipeline to aggregate and standardize over 1,500 heterogeneous SO-100 datasets from Hugging Face, creating a large-scale, unified pre-training corpus.
- Used Qwen-2.5VL-3B, an open-sourced vision-language model (VLM), to automatically annotate robot instructions to further improve data quality.
- Proposed cold-start pretraining strategy with small-scale high-quality datasets. Decreased the pretraining loss by 50% with this pretraining strategy.
- Finetuning on self-collected data from the pretrained model improves the success rate by 67% on pick-up tasks compared to direct finetuning from scratch.

Generative RNN

- Investigated how the human brain models uncertainty by modeling RNN as a generative model implementing denoising diffusion probabilistic model.
- Developed a novel computational model, the Reservoir-Sampler Network (RSN), mimicking brain-like processes for probabilistic computation. The PyTorch code is open-sourced.
- Using mathematical techniques (stochastic differential equations) to understand how neural circuits can implement score matching, and whether neural circuits are expressive enough to implement score matching. This offers insights into both neuroscience and artificial intelligence.

Sharpness and compression of neural representations

- Explored how sharpness (trace of Hessian of the loss) can affect the adversarial robustness of neural representations during training.
- Proved sharpness-related bounds on specific mathematical quantities (volume compression, maximum local sensitivity) that characterize the robustness of representations.
- Conducted practical experiments with deep learning architectures such as VGG-11, MLP, and ViT networks.
- Showed how these mathematical properties affect the network's generalization performance, crucial for improving network's adversarial robustness.

AWARDS

Boeing Research Award 2024

Violet Higgitt Frank Scholarship 2020

International Collegiate Programming Contest (ICPC), **World Finalist** 2019

Mathematical Contest in Modeling (MCM) **Meritorious Winner (Top 10% out of 10670 teams)** 2018

INVITED TALK

20 Years of Collaboration in Computational Neuroscience (University of Chicago, Grossman Center)
Talk: Expressive probabilistic sampling in recurrent neural networks (Oct. 2023)

SERVICE

Reviewer for *NeurIPS, ICLR, ICML*

Co-reviewer for *Current Opinions In Neurobiology*

TECHNICAL STRENGTHS

PyTorch, Python, MATLAB, C++, Java, R, Linux, LaTeX, and SQL

COURSE WORK

- AMATH 563 - Introduction to Deep Learning Applications and Theory
- CSE 546 - Machine Learning
- AMATH 584 - Applied Linear Algebra
- AMATH 562 - Advanced Stochastic Process
- COMP SCI 536 - Introduction to Programming Languages and Compilers
- COMP SCI 537 - Introduction to Operating Systems
- STAT 509 - Introduction to Mathematical Statistics
- CSE 571 – AI-based Mobile Robotics