Alex Wang

Summary

Stanford AI PhD developing next-generation sequence model architectures that bridge theoretical insights with practical applications. I adapt quickly across domains, from language modeling to time-series forecasting to healthcare. Every project I've led, across Citadel, AWS AI, and my own research, began in new territory and ended with practical advances. I move fast, think deeply, and do my best work when tackling open-ended problems. I'm looking for a team advancing the capabilities of next-generation AI, turning breakthroughs into products that change how people live and work.

Education

2020–2025 **PhD** | **Computer Science**, *Stanford University*, Stanford, CA, GPA: 4.0, Advised by Emily B. Fox.
I develop efficient sequence models with applications to language modeling and real-world longitudinal health data

2018–2020 **MS** | **Computer Science**, *Cornell University*, Ithaca, NY, GPA: 4.0.

2014–2018 **BA** | **Triple major: Computer Science, Math, Physics**, *Cornell University*, Ithaca, NY, GPA: 3.9.

Select publications (see Google Scholar for others)

In review **Test-time regression: a unifying framework for designing sequence models with associative memory**.
A new way of understanding sequence models, like Transformers and SSMs, and derives new model architectures, including higher-order generalizations of self-attention.

ICML 2023 **Sequence Modeling with Multiresolution Convolutional Memory**.
State-of-the-art architecture for sequences using $O(L \log L)$ memory, outperforming Transformers and S4

ML4H 2023 *Spotlight* **Interpretable Mechanistic Representations for Meal-level Glycemic Control in the Wild**.
Unsupervised, non-invasive, diabetes subtyping for noisy and missing real world CGM and meal data

ICLR 2022 **Is Importance Weighting Incompatible with Interpolating Classifiers?**.
A new loss function that makes large neural networks more robust to data distribution shifts

NeurIPS 2020 *Spotlight* **Simplifying Lagrangian and Hamiltonian Neural Networks via Explicit Constraints**.
A more accurate and data-efficient way to learn physics from dynamics data

NeurIPS 2019 **Exact Gaussian Processes on a Million Data Points**.
Make Gaussian processes usable on 100x larger datasets than before using multi-GPU acceleration

Experience

Summer 2024 **Quantitative Researcher Intern**, *Citadel GQS*, New York City, NY.
- Developed alternative alpha strategies using large-language models (LLMs); text-processing, backtesting, evals

Summer 2021 **Applied Research Scientist Intern, Forecasting**, *Amazon AWS AI*, Palo Alto, CA.
- Proposed a new approach to continuous time series forecasting using graph neural networks and neural ODEs
- Developed GPU-optimized temporal point process models for scalable training on discrete event time series
- Published findings at NeurIPS 2021 ICBINB Workshop as a spotlight paper

Summer 2018 **NLP Research Engineer Intern**, *Nuance (acquired by Microsoft for $20B)*, Greater Boston, MA.
- Built and trained Transformer-based machine translation systems for industrial-scale text datasets in TensorFlow
- Explored transfer learning strategies with unlabelled and cross-lingual data to improve model accuracy
- Integrated FastText embeddings into core internal library to improve word representation quality

Technical skills

Programming: Python, PyTorch, Jax, Pandas, Numpy, Git, Bash, Figma
AI/ML: Sequence modeling, representation learning, time-series, real-world data wrangling

Awards, fellowships, and other

2022-2024 Stanford Data Science Scholar. A prestigious PhD scholarship for $60,000 support over two years

2019 1st place team at the Cornell Data Open data science competition hosted by Citadel, $20,000 prize

alxwang@cs.stanford.edu • keawang.github.io • keawang