

SHANE CHU

+1 917-640-7301 | guixien@gmail.com | kchu25.github.io | github.com/kchu25

PROFESSIONAL SUMMARY

PhD-trained ML researcher with expertise in deep learning, mathematical optimization, and Bayesian machine learning. Achieved 150–10,000× speedups through custom CUDA kernels and optimization techniques. Specializes in interpretable ML and high-performance computing, with over six years of research experience.

PROFESSIONAL EXPERIENCE

Postdoctoral Researcher

Columbia University

Jan 2024 – Present

New York, NY

- **Deep Learning Interpretations:** Proved and validated the built-in interpretability of deep learning architectures on object-like inputs. demonstrated through applications involving high-throughput functional genomics/proteomics screenings.
- **Scaled Bayesian Optimization:** Implemented Bayesian optimization via customized CUDA kernels with Nyström approximation and randomized SVD, enabling a 10,000× speedup over CPU for oligonucleotide drug screening over 200K sequences of length 35 base-pairs.
- **Accelerated PPR–RNA Binding Site Mapping:** Reimplemented an in-house method, PPRdecoder, that maps PPR protein to its binding sites with a 150× speedup over the original implementation.
- **Epistasis modeling for protein mutagenesis:** Modeled interactive effects of mutations in LacI for large-scale mutagenesis data; integrated with wet-lab validation experiments.

Doctoral Researcher

Washington University in St. Louis

Sep 2018 – Dec 2023

St. Louis, MO

- **Sparse representations for biological sequence analysis:** Developed and quantified a sparse representation framework, packaged as MOTIFs.jl, capable of identifying complex patterns that traditional methods missed in standard databases (JASPAR, FactorBook).
- **Pattern discovery software:** Developed SeqShuffle.jl and MotifPvalue.jl, enabling statistical inference of DNA patterns across genomic datasets.

ML Research Intern

Regeneron Pharmaceuticals

May 2023 – Aug 2023

Tarrytown, NY

- **Bayesian optimization for promoter design:** Applied Bayesian machine learning and metaheuristics to design synthetic biological sequences (promoters) for AAV gene therapy.
- **Cross-species regulatory analysis:** Developed bioinformatics workflow to identify conserved regulatory elements across primate species.

Research Assistant

Center for Remote Sensing of Ice Sheets

Nov 2017 – May 2018

Lawrence, KS

- **HPC & model selection:** Built scalable optimization pipeline integrating hyperparameter tuning and model selection for ice detection from radar imagery.

TECHNICAL SKILLS

Programming & Systems

- **Languages:** Python, Julia, C/C++, R, MATLAB
- **High-Performance Computing:** CUDA, custom kernel development, distributed computing (Slurm, DDP)
- **Performance Optimization:** Custom memory allocation, randomized algorithms, CUDA kernel fusion

Machine Learning & Deep Learning

- **Frameworks:** PyTorch, Flux.jl
- **Architectures:** CNNs, Transformers, RNNs/LSTMs, GNNs
- **Advanced Techniques:** Custom backward passes, interpretability methods (integrated gradients, Shapley values)

Mathematical Methods

- **Optimization:** Bayesian optimization, operator splitting (e.g. ADMM), metaheuristics
- **Statistical modeling:** Bayesian nonparametrics, scalable Gaussian processes

- **Data Processing:** Pandas, DataFrames.jl, efficient I/O pipelines

Development & Deployment

- **Workflow:** Snakemake, Docker, Git
- **Documentation:** \LaTeX , technical writing

SELECTED WORK

- Chu, S. K., & Zhang, C. (2025). Revealing Deep Learning Interpretations for Biological Sequence Analysis and Beyond (In preparation).
- Chu, S. K., & Stormo, G. D. (2023). Finding motifs using DNA images derived from sparse representations. *Bioinformatics*, 39(6), btad378.
- Paden, J. D., Berger, V., Al-Ibadi, M., Chu, S. K., Xu, M., Crandall, D., & Fox, G. (2018, December). Subglacial bed topography using machine learning and geostatistical analysis applied to 2D and 3D radar sounding. In *AGU Fall Meeting Abstracts* (Vol. 2018, pp. C43A-07).
- Berger, V., Xu, M., Chu, S. K., Crandall, D., Paden, J., & Fox, G. C. (2018, July). Automated tracking of 2D and 3D ice radar imagery using VITERBI and TRW-S. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 4162-4165). IEEE.

INVITED TALKS

- Interpretable deep learning for protein-DNA interactions, System Biology Seminar, Columbia University Jan 2025
- DNA motifs revealed through sparse representations, System Biology Initiative, Columbia University Oct 2024
- Poster, Machine Learning in Computational and Systems Biology, Intelligent Systems for Molecular Biology Jul 2022

OPEN SOURCE CONTRIBUTIONS

EntroPlots.jl: Visualization library for complex sequence patterns.
EpicHyperSketch.jl: Parallel enumerations/counting on GPU with fixed space complexity via count-min-sketch.
AutoComputationalGraphTuning.jl: Visualization library for complex sequence patterns.
MOTIFs.jl: Advanced pattern discovery method that finds complex sequence patterns.
RealLabelNormalization.jl: Common subroutines for normalizing real-valued labels prior to ML training.
SEQ2EXPdata.jl: Data structures to handle sequence to expression (string to label) tasks.
SeqShuffle.jl: Shuffle strings such that it preserves the k-mer frequency in each string.
MotifPvalue.jl: An approximation algorithm that estimates the score-thresholds of binding sites.

EDUCATION

Washington University in St. Louis

Ph.D. in Computer Science

St. Louis, MO

Aug 2018 - Dec 2023

- Focus: Machine Learning, Optimization, Algorithm Design, Computational Biology.
- Thesis: Interpretable deep learning via sparse representations for protein-DNA interactions.
- Advisor: Gary D. Stormo

University of Kansas

Bachelor of Arts in Mathematics

Lawrence, KS

Aug. 2014 – May 2018

- Completed most computer science courses offered, including operating systems, computer architecture, data structures, and algorithms.

TEACHING & LEADERSHIP

Assistant Instructor

Washington University in St. Louis

Jan 2021 – Jun 2022

St. Louis, MO

- Designed curriculum for mathematical optimization and Bayesian ML courses (ESE 415, CSE 515T).
- Mentored 80+ students in optimization and advanced ML concepts, both theory and implementation.

LANGUAGES

- English (Fluent) and Mandarin (Native).