Zhexiang (Bruce) Li

bruceli@ucla.edu | +1 (831) 239-2728 | Los Angeles, CA | Website | GitHub | LinkedIn

Available for internships: Jun 13 – Sep 17, 2027

Education

University of California, Los Angeles Expected Jun 2028
M.S. in Computer Science

University of Southern California May 2026
B.S. in Computer Science, *Magna Cum Laude*; GPA: 3.76/4.00; Viterbi Dean's List

Research Skills

Methods: multimodal learning, efficient attention, token routing, molecular property prediction
Tools: Python, PyTorch, Triton, RDKit, Chemprop, scikit-learn, C/C++, SQL, Linux

Research Experience

Health Intelligence Lab, UCLA | *Graduate Researcher* Aug 2026 – Present
Advisor: Prof. Yuzhe Yang

- Working on multimodal foundation models for diabetes that forecast glucose from continuous glucose monitoring (CGM) together with insulin, meals, activity, and patient context.
- Harmonized CGM data from 49 datasets into a common format with meal, insulin, and wearable context; now building patient-disjoint benchmarks to compare GluFormer, T1DFormer, and custom forecasting models.

Prof. Yan Liu's Group, USC | *Drug Discovery Research* Spring 2026
Research collaboration with a Ph.D. student in the group

- Benchmarked molecular descriptor-based models, Chemprop, and hybrid approaches for SARS-CoV-2 antiviral screening across seven Therapeutics Data Commons datasets using scaffold splits.
- Analyzed descriptor importance and SHAP attributions to find the molecular features behind predictions.

Selected Papers & Research Contributions

MedTokenBudget: Lesion-Preserving Token Routing 2026
Zhexiang Li (sole author). Under review.

- Designed lesion-supervised token routing after a frozen vision backbone so that dermoscopy classifiers keep diagnostic regions under a fixed token budget; trained it with budget curriculum learning, diversity regularization, and attention distillation, and evaluated it on ISIC 2019.

Mass-Conservation Token Transformers for MS/MS Retrieval 2026
Under review at AAAI 2027

- Encoded mass-conserving fragment pairs as sparse evidence tokens for spectrum retrieval; built mass- and overlap-based hard negatives and tested how rankings change when the evidence is removed.

Efficient Attention via Pre-Scoring 2025
Zhexiang Li, Haoyu Wang, Yutong Bao, David Woodruff. Under review. arXiv:2505.11040

- Proposed pre-scoring to prioritize informative keys before approximate attention; implemented key selection in Triton as first author, advised by Prof. David P. Woodruff (CMU).
- Integrated pre-scoring with HyperAttention and evaluated it on ChatGLM with 131k-token contexts.

Additional Experience

Piezolab Medical | *AI Engineer Intern* Lake Forest, CA · Jun – Aug 2025

- Built an LLM-assisted pipeline to ingest, clean, and cluster 5,000+ medical-malpractice documents for quality-of-care research; developed agent workflows for vascular-care research.

University of California, Santa Cruz | *Machine Learning Tutor* Spring 2023

- Led office hours and co-designed Jupyter assignments on logistic regression, SVMs, and tree models; contributed to the coding autograder for Prof. Lise Getoor's course.