Alan Liang

925-428-3809 | alqliang@gmail.com | in/alnliang/ | github.com/alnliang | orcid.org/0009-0002-9620-8131

Research Interests

Biomedical machine learning, biomedical data science, molecular and cellular interaction modeling in immunology, natural language processing, representation learning, embeddings, adversarial training, tree-based methods, structure-aware protein binding prediction, uncertainty-aware and cross-reactivity-aware TCR–pMHC recognition, cancer vaccine design.

Education

University of Southern California 08/25 – 05/27 (Expected)
M.S. in Computer Science

University of California, Riverside 09/21 – 03/25
B.S. in Computer Science

Selected Work

PIT-GCL: Protein Interaction using Topological Graph Contrastive Learning
Jae Won Choi, Ryoonki Hong, Alan Liang, Manjula Adiveppa Wader, Bingsong Zeng, Peiyang Tang, Longwei Liu, Ruishan Liu
Submitted to NeurIPS 2026

A Stacking-Based Ensemble Approach for Predicting Chess Puzzle Difficulty
Alan Liang, Cenzhi Liu, Kai Wang, Ethan Liu
Accepted, presented at FedCSIS 2025; published in IEEE

Experience

Research Intern - Machine Learning for Healthcare and Biomedicine 11/25 – Present
Center on Artificial Intelligence Research for Health (AI4Health) — USC Information Sciences Institute

- Contributed to PIT-GCL, a structure-aware protein interaction model combining ESM-2 embeddings, Cα geometry, persistent homology, and cross-attention.
- Benchmarked TCR–pMHC binding models and built preprocessing and evaluation pipelines for lab-curated immunology data.
- Initiated a positive-unlabeled learning study for uncertainty-aware and cross-reactivity-aware TCR–pMHC prediction.
- Investigated cell-level biomedical interaction modeling for immunotherapy and cancer vaccine design.

Undergraduate Research Assistant - LLM Safety and Optimization Research 01/24 – 06/24
University of California, Riverside

- Investigated LLM unlearning, hallucination mitigation, adversarial robustness, and utility preservation.
- Built evaluation programs for diversity, fluency, and semantic preservation using BLEURT.
- Conducted literature review on state-of-the-art methods in LLM safety and optimization.

Projects

Cross-Reactivity-Aware TCR–pMHC Binding Prediction 06/26 – Present
Lead Project | USC Information Sciences Institute

- Conducting a feasibility study on uncertainty-aware TCR–pMHC prediction under noisy or untested negative labels.
- Investigating cross-reactivity limitations in binary binding models such as TEPCAM, STAG, and TRAP.
- Developing methods to distinguish true non-binding pairs from untested or uncertain TCR–pMHC interactions.
- Evaluating positive-unlabeled learning as a framework for handling uncertain negative labels.

Mitigating Dialect Bias in Toxicity Detection 02/26 – 05/26
Course Research Project | CSCI 567 — Machine Learning

- Implemented adversarial debiasing for toxicity detection using BERT representations.
- Designed group-specific threshold tuning and calibration to reduce AAE/SAE FPR disparities.
- Analyzed fairness-performance tradeoffs across F1, FPR, and FNR under in-domain and transfer settings.

HaluGuard: Hallucination-Aware Context Selection for Code Generation 02/26 – 05/26
Course Research Project | CSCI 544 — Natural Language Processing

- Processed RepoBench data and supported the contrastive training setup for repository-level code generation.
- Implemented retrieval baselines and evaluation using Recall@k, Exact Match, Edit Similarity, and CodeBLEU.
- Investigated hallucination-aware context selection beyond similarity-based retrieval for multi-file code completion.

Technical Skills

Languages: Python, Java, Scala, C++, JavaScript, TypeScript, SQL
Frameworks: AlphaFold, React, Node.js, Express.js, Flask, FastAPI
Libraries: PyTorch, PyTorch Geometric, XGBoost, Biopython, SciPy, Transformers (Hugging Face), scikit-learn, pandas, NumPy, TensorFlow, Matplotlib
Tools & Platforms: Git, Docker, GCP, AWS, Hadoop, Spark, Azure