Jason (Junjie) Zhu

jasonjunjiezhu.com

Email: junjie.zhu.jason@gmail.com
Mobile: 650-285-7123

Summary

I am a curiosity-driven and scientifically-trained builder with 15+ years of experience in AI/ML, Statistics, and Graph Algorithms. I have had extensive experince working with world-class scientists and top-tier product teams to create collective impact. Drawn to complex problems and high-agency teams, I build scalable solutions that leverage insights from evolving domains: multi-modal RAGs, intelligent search systems, and biomedical breakthroughs.

Education

Stanford University — Stanford, CA
Ph.D. in Electrical Engineering · M.S. in Statistics — *2014 – 2020*

Olin College of Engineering — Needham, MA
B.S. in Electrical and Computer Engineering — *2010 – 2014*

Experience

Nexa AI — Cupertino, CA
Head of AI/ML — *Feb 2025 – Present*

- Leadership: Spearheading a small team to optimize Gen-AI edge inference across CPU/NPU/GPU platforms.
- Local RAGs: Pioneering privacy-first RAG systems with lightweight AI models and edge vision processing.
- Agentic Systems: Architecting next-generation AI applications using emerging protocols (e.g., MCP, A2A).

Apple — Cupertino, CA
Machine Learning Engineer — *Jan 2020 – Feb 2025*

- Mentorship: Empowered engineers to achieve recognition through internal and external conference publications.
- Synthetic Data Generation: Pioneered robust frameworks using advanced dimensional perturbation techniques.
- Preference Learning: Engineered offline A/B testing to capture preference dynamics and distribution shifts.
- System Testing: Built high-performance Java pipelines for Maps Search query understanding evaluation.
- Ranking Triage: Created linear-time algorithms to decode multi-ranker system impacts.

Stanford University — Stanford, CA
Research Assistant — *Sep 2014 – Feb 2020*

- Graph Visualization: Engineered interactive visualizations for Gene Ontology analysis.
- Statistical Inference: Enhanced multiple-hypothesis testing to eliminate data snooping bias.
- Unsupervised Learning: Developed novel dimension-reduction techniques for cellular modeling.
- Selective Inference: Implemented advanced methods for tissue-specific eQTL analysis.
- Sequence Alignment: Enhanced DNA sequence alignment performance through C/C++ optimization.

Olin College of Engineering — Needham, MA
Research Assistant — *Sep 2010 – May 2014*

- Graph Theory: Resolved distance-2 graph coloring challenges for specialized graph families.
- Information Theory: Developed stochastic geometric models for wireless network interference.

Selected Publications

1. Automatically Authoring Regression Tests for Machine-Learning-Based Systems. *ICSE, 2021*
2. Progenitor identification and SARS-CoV-2 infection in human distal lung organoids. *Nature, 2020*
3. Exploratory gene ontology analysis with interactive visualization. *Scientific Reports, 2019*
4. Visualization and analysis of sc-RNA-seq data by kernel-based similarity learning. *Nature Methods, 2017*

Full list shown on Google Scholar: https://scholar.google.com/citations?user=2EasRdEAAAAJ&hl