Jason (Junjie) Zhu, Ph.D.

https://jasonjunjiezhu.com

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

Summary

I am a curiosity-driven, scientifically-trained builder with over 10 years of experience in AI/ML, statistics, and graph algorithms. My academic work has received over 10,000 citations, and beyond research, I have led and built production systems from the ground up: designing systems, writing production code, shaping direction, and delivering impact alongside top tier product teams. These include scalable systems in multi-modal RAGs, intelligent search, and biomedical discovery. I value clear, thoughtful communication as much as technical precision, and care deeply about getting the details right. I have found my greatest motivation comes not just from building things that work, but from helping people grow, especially those who bring diverse perspectives and backgrounds.

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*

- **RAG Productization**: Led a team of four to develop and launch an on-device RAG application in under three months, leveraging continuous regression testing and weekly iterations to accelerate quality improvements.
- **Semantic Search Innovation**: Invented, prototyped, and deployed a semantic file-search system with structured metadata, vector search, and @-search support, significantly improving file recall quality.

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

- **Offline Evaluation**: Spearheaded the development and rollout of an offline evaluation service to rigorously quantify feature impact on user experience prior to public releases, enabling faster improvement cycles. Owned multi-year roadmap strategy and stakeholder alignment. Scaled the initiative from a solo effort to a five-person team, empowering members to independently drive new scopes and sustain the roadmap beyond my tenure.
- **Research Innovation**: Designed novel generative and retrieval-based frameworks to evaluate million-scale daily traffic in Apple Maps Search; presented methodologies at top-tier software engineering conferences (*ICSE*, *FSE*).
- **Infrastructure Modernization**: Revamped internal testing pipelines for query understanding and ranking, reducing release cycles from weekly to daily and improving launch stability for WWDC-featured products.
- **Technical Leadership**: Mentored team members in defining project scopes and preparing presentations for internal AI/ML conferences, fostering professional growth and cross-team visibility.

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

- **Full-Stack Data Science**: Developed an interactive tool to visualize and perform power analysis on 30,000+ Gene Ontology terms, enabling large-scale association discovery with controlled false discovery rate.
- **Scalable Graph Learning**: Built graph-based unsupervised learning pipelines for million-scale, high-dimensional datasets, resulting in publications in *Nature*, *Nature Methods*, *Cell*, and *NeurIPS*.

Technical & Scientific Expertise

- **Core tools in regular use**: Python, Java, SQL, Pandas, NumPy, Git, CI/CD, REST APIs
- **Academic publications**: Available at: https://scholar.google.com/citations?user=2EasRdEAAAAJ&hl