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 I've had the opportunity to contribute to production-grade systems in close collaboration with top-tier product teams. I am passionate about solving complex problems and building reliable, scalable tools across emerging areas like multi modal RAGs, intelligent search, and biomedical discovery.

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 4-member team to develop and launch an on-device RAG application in under 3 months, powered by continuous regression testing and weekly iteration to accelerate quality improvements.
- **Semantic Search Innovation**: Invented, prototyped, and shipped a semantic file-search engine with structured metadata, vector search, and @-search support, improving file recall quality in customer-facing RAG applications.

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

- **Offline Evaluation Service**: Spearheaded the development and rollout of an offline evaluation framework to continuously 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 5-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; shared 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 preparing and presenting their work at internal AI/ML conferences, fostering professional growth and cross-team visibility.

Stanford University Stanford, CA
Graduate 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**: Developed graph-based unsupervised learning pipelines for million-scale, high-dimensional datasets—resulting in publications in *Nature*, *Nature Methods*, *Cell*, and *NeurIPS*.

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