Martin Kim

martinkim126@gmail.com | linkedin:martinkim0 | martinkim.me

Software engineer with a passion for cloud infrastructure, MLOps, and research engineering, with biotech experience working with both ML and lab research scientists.

Experience

Software Engineer 2024 - now

Insitro - South San Francisco, CA

- Develop & maintain custom microscopy software for data acquisition & processing for internal scientist use driving drug discovery insights (SQL, NumPy, Pandas, PyTorch, Ray, AWS Batch).
- Lead re-architecting of data extraction service processing 100 TBs of image data per month to reduce latency from hours to seconds and maintain 99% uptime (Docker, Grafana, Terraform, AWS Lambda).
- Establish org-wide pattern for deploying GPU-accelerated model inference as a service on Kubernetes (FastAPI, AWS EKS, Terraform, GitHub Actions).
- Integrate into adiposity therapeutic area as dedicated technical operations staff for ML and wetlab scientists.

Research Software Engineer 2022 - 2024

University of California, Berkeley (Yosef Lab) - Berkeley, CA

- Developed open-source package for deep probabilistic modeling of single-cell and spatial omics data with 50k+ downloads/month (PyTorch, Pyro, Lightning).
- Composed modular pipelines for data processing, model training, and reproducible analyses (Dagster, GitHub Actions, and AWS).
- Led collaboration with industry partner to fine-tune models and provide embeddings for widely-used data corpus consisting of 100M+ cells.
- Provided technical support for research scientists in model training, tracking experiments, and writing performant code.

Research Software Intern 2022

Genentech Research & Early Development - South San Francisco, CA

- Implemented a method to efficiently query a single-cell RNA-seq database of 100M+ observations for gene signatures and vector embedding similarities.
- Collaborated with research scientists in applying query methods to generate novel biological insights in immunology and oncology.

Skills

General: Python, C, SQL, GitHub Actions, Linux, Docker, Grafana
Infrastructure: Terraform, AWS, Kubernetes
Machine learning: PyTorch, JAX, Ray, Lightning, NumPy, Pandas

Education

B.S. Electrical Engineering & Computer Science 2018 - 2022
B.S. Bioengineering
University of California, Berkeley

Software

scvi-tools: Lead developer of a library for deep probabilistic analysis of single-cell and spatial omics data. 1.4k+ GitHub stars and 30k+ downloads/month. 2022 - 2024

scverse: Core developer for software consortium focused on single-cell omics analysis. 2022 - 2024

CZI CELLxGENE Census Models: Collaborating developer for model embeddings over the CELLxGENE Census data corpus. 2023 - 2024

Publications

P Boyeau, J Hong, A Gayoso, **M Kim**, JL McFaline-Figueroa, MI Jordan, E Azizi, C Ergen, N Yosef. "Deep generative modeling of sample-level heterogeneity in single-cell genomics". *Nature Methods*. 2025

C Ergen, V Valiollah Pour Amiri, **M Kim**, A Streets, A Gayoso, N Yosef. "Scvi-hub: an actionable repository for model-driven single cell analysis". *Nature Methods*. 2025

C Ergen, G Xing, C Xu, **M Kim**, M Jayasuriya, E McGeever, AO Pisco, A Streets, N Yosef. "Consensus prediction of cell type labels in single-cell data with popV". *Nature Genetics*. 2024

E Multrier, **M Kim**, P Boyeau, N Yosef, C Ergen. "Incorporating cell type ontology into probabilistic modeling for single-cell sequencing with muANVI". *Machine Learning in Computational Biology* (poster). 2023

HC Bravo, J Kancherla, G Heimberg, A Lun, V Wang, I Borozan, T Kuo, **M Kim**, S Lelong, V Obenchain. "The Single Cell Hub at Genentech, or, how to manage and query data from 100M cells". *CSHL Biological Data Science*. 2022

Grants

Chan Zuckerberg Initiative Essential Open Source Software for Science grant in support of scvi-tools (Key personnel). 2024