Kyler Zook

Bloomington, IL ● kylerzook@gmail.com ● https://www.linkedin.com/in/kylerzook/

EDUCATION

University of Missouri - Columbia | **Bachelor of Science** in Computer Science *August 2024 - Est. May 2028*
GPA: 3.84, *High Dean's List* All Semesters | **Accelerated Master of Science** in Computer Science
| Minor in Mathematics, Minor in Biological Sciences

EXPERIENCE

Machine Learning Research – Gene Regulatory Network Inference *Aug 2025 - Present*

University of Missouri - Columbia | Columbia, MO

- Architected a multi-source data engineering pipeline that ingests, normalizes, and integrates large-scale public genomics resources into a unified, versioned schema spanning 200+ cell types across human and mouse reference genomes.
- Designed a source-driven discovery and cataloging system that programmatically indexes experimental metadata via public APIs, resolves inconsistent naming across independent repositories, and ranks candidate datasets by coverage, replacing manual selection with resumable, configuration-driven batch builds.
- Developed automated quality-control tooling that benchmarks generated networks against established reference standards, quantifies coverage gaps, and diagnoses root causes of discrepancies, establishing objective acceptance gates before model training.
- Building evaluation infrastructure and baseline graph neural network models for link prediction over biological graphs, with reproducible protocols and multi-seed statistical reporting (AUROC/AUPRC); public dataset release and manuscript targeted for 2027.
- Contributed, as a first-author, to a research effort on LLM agent–driven network inference, developing an automated literature-validation agent and a multi-omic validation component integrating chromatin accessibility and transcriptomic data; **manuscript under review**.

PUBLICATIONS

Under Review

A. Hedge, **K. Zook**, R. Kim, J. Cheng. "Agent-Based Gene Regulatory Network Inference." *NeurIPS*, 2026. **Under review.**

In Preparation

K. Zook, J. Cheng. "A Benchmark for Evaluating Graph-Based Gene Regulatory Network Inference." *Manuscript in preparation*, target submission December 2026

SKILLS

Languages: *Python, C/C++, Java, SQL, JavaScript/TypeScript, Bash*
Libraries and Frameworks: *Pandas, NumPy, PyTorch, TensorFlow, SciPy, Scikit-learn, Matplotlib, Seaborn, PyArrow, SQLAlchemy, LangChain, FastAPI, pytest*
Infra/Tools: *Docker, PostgreSQL, MongoDB, Redis, Git, GitHub Actions, Linux (Ubuntu)*

PROJECTS

GRN-Bench — Distributed Benchmark Platform | *Python, FastAPI, PostgreSQL, Redis, Celery, Docker, GitHub Actions, S3*

- Self-hosted platform running multi-seed graph ML experiments as queued jobs: FastAPI + Redis/Celery worker pool with idempotent writes under at-least-once delivery; Dockerized, 74 tests in CI.
- Normalized PostgreSQL schema (datasets → versions → runs → metrics); composite indexes cut leaderboard p95 from 1.2 s to 35 ms over 10k runs.
- Content-addressed dataset versioning (SHA-256 manifests on S3) with git SHA, seed, and lockfile per run; any published number reproduces in one command.

Redis-Inspired Multithreaded Key-Value Store | *C++*

- Implemented custom memory management and efficient serialization protocols to optimize throughput and reduce heap fragmentation.
- Developed TCP networking layer and request parsing engine, enabling real-time SET/GET operations with benchmarking and stress testing for performance validation.