Stanford

Tauska Lan

Affiliate, Genetics - BASE

Bio

BIO

I'm an ML engineer specializing in LLM post-training and agentic systems—with a particular focus on domains where rigor matters: health, biology, and scientific discovery.

Long-horizon agents — Designed and shipped multi-step orchestration systems (Pantheon-CLI, OmicVerse Agent) that outperform general SWE-agent baselines on biomedical tasks. Built cross-provider query routing and sandboxed execution to keep complex workflows robust over extended interactions. My agents don't just respond—they plan, recover from failure, and complete real research pipelines end-to-end.

Agentic science — Created infrastructure where AI doesn't assist research—it conducts it. Vectorized 30 years of NHANES data; parallelized Bayesian kernel machine regression on Kubernetes; built TCGA/GEO pipelines that bridge wet-lab and dry-lab workflows. Co-developed OmicVerse, an open-source platform powering reproducible multi-omics and single-cell analyses across hundreds of studies.

Experience engineering — Scaled rubric-based reward datasets to 1M+ pairs; trained summary and chain-of-thought reward models via RLAIF/RLHF; delivered measurable benchmark lifts in health AI. I care about the full loop: data curation → reward shaping → careful ablation → verifiable outcome—no cherry-picked demos—just metrics that survive scrutiny.

Currently pursuing advanced agentic studies at Karolinska Institutet and Stanford!

Open-source: OmicVerse · Pantheon-CLI · RAG Web UI · AstrBot

If you're working on post-training at scale, scientific agents, or high-integrity data pipelines—I'm always interested in systems that move from promising results to verifiable outcomes. Let's talk.