Ryan Kim

224-517-7069 | ryankim17920@gmail.com | linkedin.com/in/ryankim17920 | github.com/RyanKim17920 | ryankim.me

Education

University of Illinois Urbana-Champaign — Urbana, IL
B.S. in Computer Science — *Expected May 2029*

William Fremd High School — Palatine, IL
GPA: 4.0/4.0 (UW), 5.207 (W) | AIME Qualifier (4x) | Simons Fellow — *Aug 2022–May 2026*

Technical Skills

Languages: Python, C/C++, Java, SQL (PostgreSQL), JavaScript, R, HTML/CSS.
ML/Data: PyTorch, verl, TensorFlow, DSPy, HuggingFace, Scikit-learn, Polars, Pandas, NumPy.
Methods: Model Evaluation, Survival Analysis, Cox Regression, Clinical NLP, Computer Vision, Self-Supervised Learning.
Systems/Tools: vLLM, SLURM, Docker, Linux, Git, Flask, FastAPI, MongoDB.

AI & ML Experience

ML Research Contributor — Apr 2026–Present
MedARC — *Remote*
- Drove **37% of the gain over baseline** on MedARC's **NanoPath** benchmark; built an **LLM autoresearch pipeline** with a dual-branch anti-overfitting architecture.
- Built an independent local pathology-model robustness pipeline similar to WAIV's work, surpassing WAIV-reported results on **7 of 8 tasks** and **improving scanner-noise robustness by 34%**.
- Post-trained **Qwen3.6-27B** with **GRPO**, cutting reasoning-token usage by **30%** without benchmark loss; served locally with **vLLM**.

Simons Summer Research Fellow (Computational Genomics) — June 2025–Aug 2026
Yurovsky Lab, Stony Brook University (Simons Foundation) — *Stony Brook, NY*
- Architected a **DNABERT-based** prognostic model using **Cox regression** on non-coding regulatory DNA mutations from **185 patients** to predict glioblastoma survival.
- Identified significant **age-related survival signals** and racial disparities in mutational scores.

Pathology AI Research Intern — Oct 2024–July 2026
AI in Medicine & Computational Biology Lab, Weill Cornell Medicine — *New York, NY*
- Conducted **foundation-model evaluation** across TCGA and internal cohorts by benchmarking **12 pathology models**.
- Found a **2× out-of-distribution error increase** (**MAE ∼0.10 → ∼0.20**) in tumor-purity prediction, highlighting clinical deployment risk.

Lead Technical Developer — Nov 2025–Apr 2026
MathLinks.org — *Remote*
- Led engineering for a **high-concurrency** competition platform serving **150,000+ users** across **11** partner organizations.
- Built **React/Node.js** authentication, submission, and **leaderboard workflows**.

Computational Oncology Research Intern — Dec 2024–Jan 2026
Advanced Computing & Oncology Lab, Memorial Sloan Kettering — *New York, NY*
- Built a **CPU-only local LLM pipeline** with LLaMA 3.1-8B, Qwen3-4B, and DSPy/MIPROv2 to classify **370 clinically annotated** lung cancer CT reports.
- Applied Cox regression and **survival analysis** to **230 patients'** clinical and demographic records to identify correlations between percent thymic tissue and NSCLC outcomes.

Selected AI Projects

speedLM | *LLM, Speculative Decoding* — July 2026–Aug 2026
- Built an **OpenAI-compatible vLLM layer** that trains speculative drafts during idle GPU time, improving accepted tokens per verifier step by **13.4%** on **Qwen3-8B**.

micro-DINOv3 | *Pure Python, Computer Vision* — Feb 2026–Apr 2026
- Implemented **DINOv3 in dependency-free Python** with hand-written **reverse-mode autograd** and a **ViT**, EMA teacher-student distillation, DINO+iBOT objectives, KoLeo regularization, axial RoPE with register tokens, and Gram anchoring.

PapersToCode | *Python, Full-Stack, Data Pipeline* — Mar 2024–Mar 2026
- Built an automated paper-to-GitHub matching **data pipeline across 500,000+ ML/AI papers**, surfacing research that still lacked public implementations.

Deep Mixture of Experts Reproduction | *PyTorch, Computer Vision* — July 2024–Nov 2024
- Recreated the main CIFAR-10 results of Deep Mixture of Experts via Shallow Embedding (**100+ citations**) in **PyTorch**, recovering its **accuracy–compute tradeoff**.