Tailai Ying

385-256-3856 | tty6@cornell.edu | tailaiying32.github.io | linkedin.com/in/tailai-ying-099041260 | github.com/tailaiying32

EDUCATION

College of Engineering, Cornell University   Ithaca, NY
B.S. in Computer Science   *Expected May 2027*

TECHNICAL SKILLS

Languages: Python, Java, C/C++, JavaScript, TypeScript, SQL, OCaml, HTML/CSS, XML
Frameworks/Libraries: React, Next.js, Node.js, Flask, Spring Boot, Tailwind CSS
Tools/Databases: Docker, SLURM, Git, GitHub Actions, Linux, Figma, Postman, PostgreSQL, SQLite, Prisma
Machine Learning: PyTorch, NumPy, scikit-learn, Pandas, Weights & Biases, Hydra, Optuna, MuJoCo, PyBullet, Gymnasium

EXPERIENCE

Research Assistant   May 2025 – Present
EmPRISE Lab   *Ithaca, NY*
- Designed and implemented the active learning core of an LLM-informed caregiving robot personalization system using Bayesian optimization and variational inference; collaborated closely with ML researchers to integrate LLM-informed priors and learned latent representations, reducing user assessment costs by over 3×. (RSS 2026)
- Scaled model training efficiency by 10x by integrating Optuna and Cornell's HPC Cluster, automating hyperparameter tuning and high-throughput experimentation.
- Scaled the MyoSuite musculoskeletal simulator into a high-throughput experimentation platform, adding modular policies, distributed execution via SLURM, experiment configuration via Hydra, and automated logging with Weights & Biases.

Software Engineering Intern   May 2025 – Aug 2025
Aria Lab   *Salt Lake City, UT*
- Built a high-fidelity 3D swarm simulation with modular controllers supporting diverse multi-agent configurations.
- Optimized simulation throughput by parallelizing agent updates, enabling the simultaneous evaluation of 100+ swarm configurations across HPC nodes.
- Engineered an automated data analysis pipeline using novelty search and clustering to programmatically discover and classify 10+ previously unknown agent behaviors.

Technical Lead   Jan 2025 – May 2025
CommuniCare   *Ithaca, NY*
- Led frontend development for React-based healthcare resource platform serving underserved communities, implementing reusable components and responsive design.
- Collaborated with a cross-functional team to implement Agile workflows and CI/CD, reducing SDLC time by 50%.

SELECTED PROJECTS

Jarvis - Multi-modal AI Voice Assistant | *C++, llama.cpp, sherpa-onnx*   Dec 2025 – Jan 2026
- Prototyped a high-fidelity audio-first AI assistant using Qwen3, Whisper, and Sherpa-ONNX, optimizing for natural speech flow and low-latency user interaction.
- Engineered an intelligent phrase-boundary detection system to optimize real-time inference, reducing end-to-end latency to sub-1s to ensure a seamless consumer experience.
- Orchestrated concurrent model execution and audio pre-buffering to maintain pipeline performance and scalability.
- Implemented a Retrieval-Augmented Generation (RAG) pipeline to reduce hallucinations and provide context-aware responses by retrieving domain-specific knowledge in real-time.

CritterEvo - Artificial Life Simulator | *Java*   Dec 2024 – Feb 2025
- Engineered evolutionary ecosystem simulator in Java, featuring custom AI controllers and procedural generation.
- Optimized performance with multi-threading and lazy loading, increasing simulation throughput by over 80%.
- Optimized entity navigation by implementing an A* pathfinding engine with state caching, significantly reducing computational overhead during high-density simulations.
- Achieved 95% test coverage with comprehensive JUnit suite, ensuring robust functionality across edge cases.

Ear Trainer - RCM-based Ear Training | *Next.js, Spring Boot, PostgreSQL*   Jun 2025 – Aug 2025
- Developed ear training platform featuring dynamic exercise generation and real-time music notation rendering.
- Migrated the backend architecture from TypeScript to Spring Boot, optimizing database queries and exercise generation logic to improve system response times by 30%.
- Designed SQL schema to track user progress across 10 skill levels, ensuring data integrity and efficient retrieval.

Lockd - Smart Lock System | *React Native, Flask*   Oct 2024
- Built IoT smart lock system, featuring vibration/sound anomaly detection with under 200 ms alert latency from Raspberry Pi to client, and real-time push and email notifications.
- Selected as Finalist and Beginner's Prize winner from a field of 40+ teams and 135+ hackers at BigRedHacks for technical implementation and system design.