YIN TANG

jackeyty@berkeley.edu | +1(510)703-3250 | Philadelphia, PA, 19103
https://github.com/JackeyTY | https://jackeytang.com | www.linkedin.com/in/yin-tang-jackeyty

EDUCATION

University of Pennsylvania, Philadelphia, PA — May 2025
Master of Science in Engineering, Computer Graphics and Game Technology — **GPA: 3.85**
- Relevant Coursework: GPU Programming and Architecture, Deep Generative Models, Computer Vision and Machine Perception, Computer Animation, Game Design Practicum, 3D Computer Modeling, Cryptography, Engineering Entrepreneurship

University of California, Berkeley, Berkeley, CA — Aug 2021
Bachelor of Arts, Computer Science & Applied Mathematics — **GPA: 3.65**
- Relevant Coursework: Data Structures, Algorithms, Artificial Intelligence, Machine Learning, Database Systems, Computer Graphics, Machine Structures, Optimizations, Linear Algebra, Probability Theory, Calculus

TECHNICAL SKILLS

Programming Languages: Python, C++, C#, CUDA, Java, JavaScript, SQL, Bash, GLSL
Frameworks & Libraries: PyTorch, Transformers, SciPy, NumPy, Pandas, Spring, Qt, OpenGL, WebGPU
DevOps & Cloud: Docker, Kubernetes, Git, Conda, MySQL, Navicat, AWS, Alibaba Cloud
Tools & Platforms: Unreal Engine, Unity, Maya, MATLAB, Linux, Jira, Miro, Microsoft Office

WORK EXPERIENCE

Wharton Research Data Services — Philadelphia, PA
Software Engineer, ML Infrastructure — Sep 2024 - Present
- Deployed Qwen2.5 language models on NVIDIA Triton Inference Server via Kubernetes, optimizing inference performance through monitoring, benchmarking, and analysis
- Constructed an evaluation dataset for training and assessing large language models in financial analysis by extracting and processing data from earnings call Q&A sessions, financial statements, and relevant financial news

PlusAI — Santa Clara, CA
Software Engineer Intern — May 2024 - Aug 2024
- Integrated vector database Milvus into Imgsearch infrastructure with compression index IVF_SQ8 and memory mapping, reduced both query latency and runtime memory usage by 10x compared to production Imgsearch
- Deployed Florence-2-large and Stella models in Imgsearch for image-caption embedding generation and querying, achieving more relevant and precise search results, and created a PoC for video captioning using the LLaVA model
- Designed and implemented an Airflow indexing job to process historical and daily image data into embeddings for all models on GPUs, supporting independent progress tracking for each model, progress resumption, and duplicate prevention
- Developed a truck HMI system using Unreal Engine, incorporating real-time data visualization and advanced interactive features

Bobcatminer, Inc. — New York, NY
Product Manager — Oct 2021 - Jun 2023
- Managed a team of 3 to supervise the manufacture and maintenance of 400K hotspots and complete 75+ over-the-air updates for system optimization and bug fixing
- Directed the development of 3 products by establishing critical testing milestones and timelines for prototype deliverables, authored PRDs and application testing reports, and coordinated with Industrial Design firm to conceptualize 2 casing designs
- Led the support engineering team to handle remote technical customer supports and return merchandise authorization (RMA), reducing the monthly RMAs by 85% through collaborating with primary suppliers to develop the firmware flashing tool
- Designed workflows for customer support, RMAs, and internal hotspot data registration, programmed scripts for data registration

ZonIoT — Shenzhen, China
Software Engineer Intern — Jun 2019 - Aug 2019
- Implemented new features independently for bespoke LoRa water meter management platforms using Spring Boot, and conducted interface testing to ensure seamless cross-platform interaction
- Restructured database schemas and optimized SQL queries, achieving a 25% reduction in REST API call response times

ACADEMIC PROJECTS

CUDA Path Tracer: CUDA, C++
- Developed a high-performance path tracer utilizing material sorting and termination compaction to improve kernel efficiency
- Enhanced with BVH GPU traversal, complex model loading in OBJ format, multiple importance sampling, and depth of field

Mini-Minecraft: C++, GLSL, OpenGL, Qt
- Constructed terrain biomes and cave systems through procedural generation with noise functions and implemented day-night cycle
- Created dynamic real-time shadows utilizing shadow mapping and post-process shaders for underwater and lava effect