WINSTON TSUI

New York, NY • wt285@cornell.edu • (929) 410-1689 • linkedin.com/in/winstontsui • github.com/winstontsui

EDUCATION

Cornell Tech (Cornell University)– New York, NY August 2024 – May 2025
Master of Engineering in Computer Science
Relevant coursework: Applied ML Engineering, NLP, Computer Vision, Security and Privacy, AI Compute, Data Science, HCI and Design.

Syracuse University– Syracuse, NY August 2021 – May 2024
Bachelor of Science in Computer Science, GPA: 3.92 | Graduated Summa Cum Laude | Tau Beta Pi (Top 8% Engineering GPA)
Relevant coursework: Graphics, Operating Systems, Database Management System, Software Engineering, Virtual Reality, Algorithms.

SKILLS

- **Programming Languages**: Java, Python, C, C++, C#, Kotlin, Swift, JavaScript (TypeScript, HTML, CSS)
- **Frameworks & Libraries**: **Frontend** (React.js, Vue.js, Angular, SwiftUI), **Backend** (Python, Django, Node.js, .Net), **Machine Learning** (PyTorch, TensorFlow, Scikit-learn, Numpy, Pandas, Matplotlib, Jupyter), **Database** (MongoDB, SQL Server, Postgresql, AWS, Azure)
- **Development Tools:** Jenkins, Jira, Maven, Postman, CI/CD, DevOps, Junit, Microservices, REST APIs, Apollo GraphQL, Docker, Kubernetes, Scrum Master.

EXPERIENCE

Palapa.ai - Software Engineer Intern | New York, NY May 2023 – August 2023

- Improved object detection performance by 10% in the Flutter Android app, integrating the YOLOv8 AI vision model and ONNX framework.
- Reduced app size on the Google Play Store by 50% through code optimization, reaching 500 daily active users within 6 weeks.
- Refactored backend screen navigation routes using Dart extension methods, increasing the app's MVC architecture stability.
- Collaborated with cross-functional UI/UX and marketing teams to prototype and test features, cutting product iteration cycles by 50%.
- Created CI workflows using GitHub Actions, which reduced deployment times by 40% and allowed the team to deliver features quicker.

Cornell University - Health Tech Research Assistant | New York, NY June 2024 – December 2024

- Implemented iOS background task management system using Swift and BackgroundTasks, increasing users' task completion rates by 20%.
- Added iOS notification features using Swift UserNotifications which increased the number of visits to our app by 30%.
- Led the frontend development of a data extraction feature related to RAG pipeline using leading frameworks like Vue.js and ElementUI.
- Fine-tuned our NLP health chatbot by adding significantly more data, which improved the bot's accuracy by 20%.

NYC Department of Design and Construction - Application Development Intern | New York, NY June 2022 – August 2022

- Automated the process for generating thousands of business card prototypes using Adobe InDesign, cutting time to production by 50%.
- Increased the accuracy of employee data by 20% by identifying null data bottlenecks in Excel database, improving data quality.

PROJECTS

Uber Clone Mobile App (GitHub) November 2024 - Present

- Built a cross-platform Uber clone app using React Native and NativeWind with ride-booking, authentication and payment functionalities.
- Implemented real-time user authentication using Google OAuth and Clerk which reduced onboarding time by 25%.
- Implemented real payment information using Stripe, allowing users to book nearby uber drivers, all stored in a Neon postgresql database.
- Integrated real-time location tracking, direction and map services using Google Places Autocomplete and Map Directions API.

Multi-Agent Chat App with LangFlow and Retrieval-Augmented Generation (RAG) (GitHub) November 2024 - January 2025

- Deployed a multi-agent system using LangFlow and AstraDB for real-time order and product data retrieval, ensuring fast database queries.
- Integrated custom workflows with OpenAI agents, RAG pipelines and vector database to synthesize data from AstraDB collections.
- Resolved complex integration challenges around schema mismatches and query performance–reduced query latency by 25%.
- Built a responsive Streamlit front-end which reduced end-user query times by 15%.

MiniTorch: Python Re-implementation of PyTorch API (GitHub) August 2024 - January 2025

- Developed a 100% PyTorch-compatible deep learning library with native PyTorch code.
- Designed custom Tensor data structures supporting back-end mathematics operations like broadcasting, auto-differentiation and backpropagation for model training. Engineered automated testing to verify source code, using streamlit for advanced analytics.
- Optimized deep learning training speed by 10x using parallelized map, zip, and reduce operations with Numba JIT and CUDA.

Amazon Clone Web App (GitHub) May 2024 - August 2024

- Created an open-source, full-stack app with 100+ products and loads in 40% less time with Next.js and MongoDB caching.
- Designed a scalable backend with Prisma ORM able to handle 1,000+ concurrent requests at a time.

GuacaGoalie Android App (GitHub) January 2024 - May 2024

- Directed a team of 4 developers to design and develop a step counter-based fitness app using accelerometer, gyroscope and GPS.
- Implemented SQLite for offline data storage and Google Play Services API using Java and Android Studio, supporting 100+ active users.