James Chen

(770) 309-2207 jameschen0831@gmail.com linkedin.com/in/jameschen0831

Education

Georgia Institute of Technology **Graduating December 2026**
M.S. Computer Science, Concentration: Machine Learning **GPA: 3.96**
B.S. Computer Science, Concentration: Intelligence and Infoworks

Experience

Amazon May 2025 - August 2025
Software Development Engineer Intern *Sunnyvale, CA*

- Architected ML-powered event monitoring system processing **14.4TB daily data**, serving **50+ developers**
- Developed multi-source data ingestion pipeline processing **50K+ documents** across Quip, Wiki, and dashboards for **RAG-based AI assistant** integration
- Developed real-time **Grafana dashboards** pipeline, decreasing report generation time from **10+ hours → 15 minutes**
- Built data pipeline with **99.9% reliability** using AWS CDK and CI/CD, processing **100GB+ event data per 10-minute interval**

Transformation Continuum – Google Cloud Commission November 2023 - May 2024
Full Stack Software Engineer Intern *Austin, TX*

- Delivered novel sales tooling to be used by **Google employees** using ReactJS
- Developed Node.js REST APIs, improving system integration efficiency by **40%**
- Implemented Quantitative Assessment Module with **500+ standardized questions** using React components
- Streamlined training by deploying **LLM-powered chatbot** to website for reviewing performance

Research

Georgia Tech LIDAR Lab January 2024 - Present
ML Research for Robotics under Dr. Ye Zhao *Atlanta, GA*

- Conducted research on **bipedal navigation** in human-crowded environments, addressing challenges of nonlinear dynamics in locomotion
- Developed K-means clustering algorithm and single shot ego agent, reducing processing time by **15%**
- Delivered re-targeting pipeline for humanoid datasets with **100K+ motion frames** (MoCap, Mujuco, AMASS) for RL training

Neural Data Science Lab at Georgia Tech June 2021 - April 2022
Research under Eva L. Dyer *Atlanta, GA*

- Conducted pioneering research using deep learning to study brain structures and cell types, winning the **APA Award for Outstanding Research in Psychological Science**
- Collaborated with the **Allen Brain Institute** on novel 3D brain imaging data sets, using CNN models to predict and classify signs of neurodegeneration with **98% accuracy**

Projects

Globally Shared KV Cache for LLMs | *Python, Golang, HuggingFace, LLaMA 2, AWS, Redis* May 2025

- Developed **3x cache-hit rate**, **30% memory usage reduction**, and **2.8x speedup** in inference latency compared to no-cache
- Integrated semantic similarity logic using cosine distance and embeddings for reusable queries
- Achieved **∼900ms reduction** in response time on only **∼0.5MB memory overhead**
- Simulated on **1000+ queries** across 3 datasets (real + synthetic)

Sano Gym Social Media App | *Swift (Frontend), Python (Backend), SwiftUI, Flask* August 2024

- Developed a **full-stack social media app** tailored for fitness enthusiasts, with Swift-based frontend and Python Flask backend, achieving over **1,000 downloads** within the first month
- Integrated **workout tracking**, routine sharing, and **community engagement** features to enhance user interaction and motivation

HackGT Contractly AI | *Python, Flask, React, MongoDB, LangChain* October 2023

- Built a **full-stack contract analysis tool** with LLMs, trained on **500+ contracts** spanning NDAs, mergers, and acquisitions
- Implemented a **Chrome Extension** to automatically fetch and rate websites' privacy policies using ChatGPT-3

Technical Skills and Awards

Programming Languages: Java, C, Python, C++, SQL, JavaScript, TypeScript, Swift, HTML/CSS, Dart **Libraries**: pandas, NumPy, Matplotlib, React, Angular, PyTorch, TensorFlow, Flask, Django, Spring, OpenCV
Technologies: AWS, Docker, Kubernetes, Git, REST APIs, SQL, NoSQL, MongoDB, Firebase, CI/CD, Agile, Scrum
Awards: Highest Honors, Dean's List, Faculty Honors, AIME qualifier, National Merit Finalist