Krishna Panchap

kpanchap@uw.edu | in/kpanchap | github/krishna-panchap | (425) 773-2099

Education

University of Washington Seattle, WA

BS in Computer Science and BS in Economics. Robinson Scholar. 3.96 GPA *Expected Grad. June 2026*

- Relevant Coursework: Computer Systems, Data Structures, Algorithms, Probability and Statistics, Object Oriented Design, Machine Learning, Networks, Distributed Systems, Deep Learning, Web Browser Engineering
- Teaching Assistant: CSE 311: Foundations of Computing (2x), CSE 121: Intro to Computer Programming

Experience

Software Engineering Intern June 2024 - September 2024

Amazon Shipping ($AMZN) *Bellevue, WA*

- Developed and deployed a full-stack dashboard for shippers, providing real-time status updates and AI-driven insights. Built using React, AWS API Gateway, Lambda, Bedrock, and DynamoDB.
- Architected the entire project from product requirements to design, development, unit-testing, and deployment.
- Secured Principal Engineer support and presented complete project demo to Amazon Shipping's Technical Director.
- Projected to reduce Shipper Escalation metric by 30%, saving over 200 annual shipper support hours.

AI/ML Applied Scientist Intern March 2024 – June 2024

Tenstorrent (Series D AI Chip Startup) *Santa Clara, CA*

- Led a joint-effort with Chipstack.ai to deploy their AI-assisted chip verification service on Tenstorrent servers.
- Built and deployed an inference server on Tenstorrent hardware (Wormhole) fine-tuned on the internal codebase with on-prem RAG using ChromaDB.
- Grew system adoption to the entire RISC-V development team, with 5000+ inference requests to date.
- Refactored a deprecated CLI tool (Docker + SLURM) which scheduled developer sessions across internal compute hardware to become the official tool with 40+ new users.

Machine Learning, Artificial Intelligence Engineering Intern September 2023 – February 2024

Arcade.ai (Series A AI Startup) *San Francisco, CA - Remote*

- Reduced average completion-time by 5 seconds (33%) by refactoring search-intent extraction of the AI pipeline.
- Fined-tuned numerous text-to-image models (SDXL) to improve manufacturability score by 30%.
- Utilized context via a VectorDB to assist in image generation, improving accuracy-to-prompt metric by 50%.
- Achieved a 2x decrease in time-to-completion by building an auto-scaling GPU orchestration layer with Kubernetes

Software Engineering Intern June 2023 – September 2023

Heretic Ventures (Venture Studio) *San Francisco, CA*

- Refactored 30+ Python scripts into a full-stack Django REST service MVP in a 2 week timeline.
- Created an automated data-processing pipeline with InSPyReNet using web scraping methods, shortening model training by 25% across the portfolio companies and eliminating 20+ hours of manual labor for developers.

Machine Learning Research Assistant March 2022 – April 2023

Social Futures Lab, Ubiquitous Computing Lab *Paul G. Allen School for Computer Science*

- Built a smartphone-based service for predicting stroke risk from facial blood flow patterns.
- Developed facial recognition API using OpenCV2 to improve app accuracy across skin tones by 20%.

Projects

Chapterit | *Python (Flask), JavaScript, PostgreSQL* | 400+ users October 2023 – December 2023

- Developed a live AI-powered full-stack web app to automatically chapter videos using a custom fine-tuned GPT 3.5-Turbo prompt layer.

Mezzo | *React Native, Java (Spring), Azure* | 40 active users and 100+ events planned August 2021 – June 2023

- Architected and built out 3-tier software architecture (React Native, Java-Spring) with 80k+ lines of code.

Technical Skills

Languages: Java, Python, C/C++, SQL, JavaScript, TypeScript, HTML, CSS, Swift, Excel
Frameworks: React, React Native, Node.js, Flask, Django, FastAPI, JUnit, Next.js, Express.js, SwiftUI
Developer Tools: Docker/Kubernetes, Azure, AWS Lambda /S3, Firebase, Stable Diffusion, Dreambooth, GPT, Pinecone
Libraries: Pandas, NumPy, Matplotlib, TensorFlow, PyTorch, Keras, Selenium/BS4