Chris Woo

chris.woo1348@gmail.com | github.com/ChrisWoo0443 Millbrae, California 94030

EDUCATION

University of California, Riverside — Sept. 2022 – June 2026
Bachelor of Science in Computer Science with Business Applications — *Riverside, CA*

EXPERIENCE

Outamation Extern — Jan. 2026 – Present
Extern — *Remote*
- Built modular AI pipelines for OCR, PDF parsing, and RAG-based extraction to process and search 200+ page mortgage documents using Tesseract, PaddleOCR, and PyMuPDF.
- Designed and optimized a retrieval-augmented document search system using LlamaIndex and open-source LLMs (Mistral, Phi-2), improving precision through chunk tuning, metadata filtering, and multi-document indexing strategies.
- Led end-to-end system evaluation and benchmarking, measuring OCR accuracy, RAG retrieval quality, and routing performance; delivered technical recommendations, optimization strategies, and deployment guidance, and built a demo UI for stakeholder validation.

AI Systems Engineer Intern — Dec. 2025 – Present
Deepiri — *Remote*
- Redesigned and implemented a modern testing UI using React + Vite, improving developer productivity and enabling faster validation of AI system behaviors.
- Integrated vector store management tooling, including document ingestion/removal and collection visualization, to support scalable retrieval workflows for AI applications.
- Conducted quality assurance and code reviews on pull requests, enforcing engineering best practices, improving system reliability, and reducing regressions across the AI platform.

AI Researcher — June 2025 – Present
Algoverse — *Remote*
- Currently developing a regularization technique for transformer models using mutual information (MI) to reduce redundancy between layers and improve model generalization.
- Led experimental evaluations comparing baseline and MI-regularized transformer models, analyzing the impact of layer targeting strategies on performance.
- Designed and iteratively tested fine-tuning pipelines for DistilBERT, analyzing results across accuracy, convergence behavior, and inference efficiency metrics.

PROJECTS

AIRA - AI Research Assistant | *Python, FastAPI, React, SQL, Vite, ChromaDB* — February 2026 – Present
- Designed a full-stack RAG application using FastAPI and React, enabling semantic search and conversational querying over uploaded PDF and text documents.
- Implemented a vector database pipeline with ChromaDB and Ollama embeddings for semantic document retrieval and context-aware LLM responses.
- Built real-time streaming chat interface with Server-Sent Events and SQLite session management, supporting multi-turn conversations with persistent history.

JobGPT | *Python, FastAPI, React, JavaScript, HTML, CSS* — April 2025 – May 2025
- Implemented a keyword-based matching algorithm to compute job–candidate compatibility scores and generate actionable resume improvement suggestions.
- Achieved 90%+ automated test coverage using Pytest and Vite Test, ensuring reliability and maintainability across frontend and backend components.
- Managed development workflows with GitHub Projects and implemented CI/CD pipelines using GitHub Actions for automated testing and deployment.

TECHNICAL SKILLS

Languages: Python, C++, SQL, Java, JavaScript, Typescript
AI/ML: PyTorch, TensorFlow, Transformers, scikit-learn, LLM Training Pipelines, RAG, Vector Databases
Backend, Data & Infra: FastAPI, REST APIs, PostgreSQL, Redis, Docker, Linux, Milvus, ChromaDB
Tools: Pandas, NumPy, OpenCV, Tesseract, PaddleOCR, PyMuPDF, React, Pytest, Vitest