

YASH DUSANE

+1-657-445-7419 | yash_dusane@csu.fullerton.edu | [LinkedIn](#) | [GitHub](#)

EDUCATION

Master of Science, Computer Science, California State University Fullerton
Bachelor of Engineering, Computer Engineering, Savitribai Phule Pune University

Aug 2024 — May 2026
Aug 2020 — May 2024

SKILLS

- Programming & Backend:** Python, FastAPI, Java, C++, SQL/NoSQL (PostgreSQL, MongoDB, Firebase), RESTful APIs
- AI/ML:** PyTorch, TensorFlow, Scikit-learn, Hugging Face Transformers, NLP (Classification, NER, Summarization), Computer Vision
- Generative AI:** RAG Systems, LangChain, LangGraph, Prompt Engineering, Vector Databases (ChromaDB, Pinecone), Fine-Tuning, Embeddings (MiniLM, Sentence Transformers), LLMs (GPT-4, LLaMA 3.1, Gemini Pro, Groq)
- Cloud:** AWS (ECS, S3, Lambda), Azure (Data Factory, Databricks, Synapse), Docker, Redis, CI/CD (GitHub Actions)
- Frontend & Web:** JavaScript, Next.js, React.js, HTML/CSS, Tailwind CSS, Server-Side Rendering

EXPERIENCE

Graduate Research Assistant, California State University Fullerton **Aug 2025 --- Present**

- Architected agentic AI system using LangChain and LangGraph for ESG compliance automation, processing 1,000+ filings across GRI, SASB, and TCFD standards with 85% regulatory interpretation accuracy, reducing manual review time by 40%.
- Designed RAG pipeline with ChromaDB and all-MiniLM-L6-v2 embeddings achieving <300ms query latency, implementing hybrid search for real-time Q&A over regulatory documents with improved retrieval precision.
- Implemented multi-agent workflow with specialized agents for extraction, validation, and generation, enabling parallel processing that improved end-to-end efficiency by 40% over sequential approaches.
- Evaluated prompt engineering techniques (chain-of-thought, few-shot, RAG), achieving 15% accuracy improvement through fine-tuned templates validated against benchmark datasets. Integrated 20+ domain-specific sustainability metrics through cross-departmental collaboration.

Software Developer Intern, Pesh Group

Jan 2024 --- Jun 2024

- Built production RAG-based support chatbot using Groq LLaMA API and LangChain, automating 80% of customer queries with <2s response time, reducing support ticket volume and improving satisfaction scores by 50%.
- Developed Next.js web application with SSR and dynamic routing, reducing page load times by 45% and supporting 5K+ concurrent users. Documented 15+ APIs with Swagger, cutting onboarding time by 30%.
- Optimized front-end performance and SEO through code splitting and lazy loading, achieving 35% bounce rate reduction, 2x session duration increase, and 60% organic traffic growth within 3 months.
- Implemented CI/CD pipeline with GitHub Actions for automated testing and deployment, conducting peer code reviews and maintaining 95%+ test coverage across bi-weekly Agile release cycles.

PROJECTS

Crush My Course - AI-Powered Academic Course Tracker [[Demo](#)]

- Built RAG-based chatbot using FastAPI, ChromaDB, and Groq LLaMA 3.1 (70B) processing 10K+ CSUF course records with all-MiniLM-L6-v2 embeddings, achieving <500ms semantic search latency for natural language queries on professor ratings and course selection.
- Containerized RAG pipeline using Docker multi-stage builds and deployed on AWS ECS with Fargate, implementing Redis-based query caching that reduced duplicate LLM API calls by 40%, lowering inference costs while maintaining conversation context across container restarts in auto-scaled ECS cluster.

ElderEase - AI-Powered Digital Companion for Elderly Wellness [[Github](#)]

- Developed conversational AI assistant using Gemini Pro API with RAG and prompt engineering, delivering personalized Medicare recommendations and medication management through multi-turn dialogue processing user health history.
- Deployed multi-container application using Docker with separate services for Next.js frontend, FastAPI backend, and Appwrite, orchestrating inter-container communication through Docker networks and implementing health checks for zero-downtime deployments on cloud infrastructure.
- Integrated multi-modal AI capabilities combining Gemini Vision API for pill identification, text-to-speech, and sentiment analysis to provide empathetic, context-aware support for administrative and emotional wellness needs.

Automated Monitoring of Gym Exercises through Human Pose Analysis [[Github](#)]

- Deployed real-time pose estimation system using MediaPipe Holistic (33-point tracking) and custom CNN classifier trained on 2,000+ annotated videos, achieving 92% accuracy across 5 exercise types with 30 FPS processing and biomechanical feature extraction to detect improper form.
- Built React.js web application with TensorFlow for deep learning inference, enabling browser-based pose analysis with rep counting, workout tracking, and visual overlay feedback for real-time corrective guidance.