TANUJ PANKAJ SANSARE

sansaretanuj@gmail.com +91 94041 13255 linkedin.com/in/tps2000

Education

College of Engineering Pune (COEP), India *2020 – 2024*
B.Tech. in Electronics & Telecommunications Engineering — Minor in Computer Engineering *CGPA: 8.69 / 10.0*
Coursework: Data Structures & Algorithms, Data Communication and Networking, Database Management Systems, Linear Algebra, Machine Learning, Deep Learning, Probability & Statistics
Leadership: Co-Founder Innovation Cell Initiative at COEP; Innovation & IPR Head, Pune Startup Fest 2022
Research Member: COEP Satellite Initiative (CSAT) — nanosatellite communications (ISRO-reviewed)

Experience

Bajaj Finance Limited — AI Engineer *July 2024 – Present*

- Architected an end-to-end **LLM-driven content generation system** comprising prompt orchestration, vision-based validation, and asynchronous inference services for large-scale marketing automation.
- Designed distributed worker-based inference pipelines handling **9,000+ daily generation requests**, addressing backpressure, retries, and partial failures under burst traffic.
- Re-architected the image generation pipeline by integrating optimized U^2-Net segmentation and batched inference, reducing end-to-end latency from **54 minutes to 1 minute**.
- Built a **ViT-based quality control service** producing confidence scores, enabling automated publish/reject decisions and human-in-the-loop escalation.
- Implemented metadata extraction and retrieval-augmented prompting to enforce brand, layout, and compliance constraints, improving output consistency and reducing creative rejection from **40% to 8%**.
- Applied model optimization techniques (quantization, ONNX Runtime) to improve throughput and system stability.

Bajaj Finance Limited — Software Development Intern *Jan 2024 – Jun 2024*

- Developed a full-stack **KPI analytics platform** with React-based dashboards and RESTful backend services supporting **16+ enterprise metrics**.
- Engineered backend data pipelines for metric ingestion, aggregation, and persistence, enabling low-latency analytical queries for business and engineering teams.
- Performed Core Web Vitals optimization on high-traffic pages, improving page load times by **15%**.
- Collaborated with cross-functional stakeholders to translate ambiguous business requirements into stable, production-ready software components.

Samsung Research (PRISM) — Machine Learning Research Intern *Aug 2023 – Jan 2024*

- Developed a multimodal sentiment analysis system combining vision-based region segmentation and transformer-based language models for customer chat screenshots.
- Modeled spatial layout information to preserve conversational structure absent in text-only approaches, improving robustness to visual variation.
- Trained and evaluated models on **8,000+ annotated samples**, achieving a **0.84 F1-score**.

American Axle & Manufacturing — Intern *May 2023 – July 2023*

- Built a **semantic retrieval system** over large-scale engineering documentation using dense embeddings and FAISS-based similarity search.
- Designed hierarchical chunking and embedding deduplication strategies, reducing vector index memory usage by **85%** while preserving retrieval quality.
- Optimized ingestion and indexing pipelines for on-premise deployment under constrained compute and memory environments.

Technical Skills

Programming: Python, TypeScript/JavaScript, C/C++, SQL
AI/ML: PyTorch, Vision Transformers, U^2-Net, RAG Pipelines, Langchain, ONNX Runtime
Backend & Infrastructure: Node.js, PostgreSQL, Redis, RabbitMQ, Celery, Docker, AWS (EC2, S3), FAISS, VertexAI

Honors & Awards

Gold Medal — NPTEL Machine Learning for Engineering and Science Applications (Top 1%, IIT Madras, 2024)
Silver Medal — NPTEL Introduction to Database Systems (Top 1%, IIT Madras, 2024)
Publication — Co-Authored "Delineation of a Robust Communication System and Protocol for Nanosatellites," & Presented at *73rd International Astronautical Congress (IAC)*, Paris, France, 2022