

SHUBHANKAR JOSHI

San Jose, CA | 541-250-1125 | joshishu16@gmail.com | [linkedin.com/in/s-joshi16/](https://www.linkedin.com/in/s-joshi16/) | shubhu163.github.io/sj/

SUMMARY

Software Engineer with expertise in software development, AI, system design, Python, and SQL, seeking opportunities to deliver scalable and intelligent solutions.

EDUCATION

Master's in Computer Science – Oregon State University – **GPA: 4.0 / 4.0** **Sept 2023 - Dec 2025**

Coursework: Algorithms, Machine Learning, Deep Learning, Data Visualization, NLP, HCI, System Design

Bachelor's in Instrumentation & Control – University of Pune – **GPA: 3.9 / 4.0** **Aug 2019 - May 2023**

Coursework: Data Structures, Operating Systems, Computer Architecture, Database Systems, Computer Networks.

SKILLS

AI & ML: PyTorch, Scikit-learn, Hugging Face, LLMs, LangChain, Vector Databases, Predictive Modeling.

Software Development: Python, C++, JavaScript, TypeScript, SQL, React, Next.js, Node.js, FastAPI, REST APIs,

Cloud: AWS, (S3, EC2, Lambda), Docker, GitHub Actions (CI/CD), Model Monitoring, Serverless Inference.

Data Systems: PostgreSQL, MongoDB, MySQL, Redis, Apache Spark, Tableau, Power BI, Statistical Analysis, Airflow.

WORK EXPERIENCE

Research Software Engineer, Hendrix Lab, Oregon State University **Sept 2023 – Dec 2025**

- Reduced experiment triage time by 40% by developing a full-stack **React, FastAPI** dashboard with **PostgreSQL**.
- Cut data processing time by 20% by refactoring ingestion into modular **Python** and **Airflow** with for 45K+ transcripts.
- Optimized GPU memory utilization by implementing batching, caching, and CI via **GitHub Actions**.
- Achieved a 0.72 correlation in predicting 100+ RNA–protein bindings by architecting Transformer model via PyTorch.
- Boosted model accuracy by 5 % through Optuna hyperparameter search and tracked experiments on Weights & Biases

Data Scientist Intern, Landmark TechEdge Pvt Ltd **Aug 2022 – May 2023**

- Enhanced pipeline efficiency by 15% by engineering scheduled data services in SQL Server and **Azure Data Factory**.
- Accelerated reporting queries by 5% for high-volume datasets through index tuning and **SQL** joins optimization.
- Achieved environment parity across Dev and Prod by containerizing batch workloads with **Docker**.
- Delivered reliable in-house tools by collaborating cross-functionally with engineering and finance teams.

Software Engineer, Samarth Enterprises **Nov 2020 – May 2022**

- Decreased manual processing time by 30% by building **Python** automation for **SQLite** ingestion and reporting.
- Scaled production ramp-ups by leading a team of 10+ technicians and managing Piaggio delivery timelines.
- Elevated defect detection by 10% using real-time **Power BI** dashboards for QC on wire harnesses and regulators

PROJECTS

InferenceHub – SaaS Payment & Subscription Infrastructure [GitHub Link](#)

- Implemented authenticated user dashboard with protected routes using Next.js, PostgreSQL, and Prisma
- Integrated Stripe subscription lifecycle including checkout, billing portal, and webhook handling for role-based access.

Khabar AI – LLM-Powered News Aggregation Pipeline [GitHub Link](#)

- Engineered a multi-threaded ETL pipeline in Python to ingest real-time data from News, Finance, and Weather APIs.
- Designed a two-stage LLM architecture (Llama 3.3 and GPT OSS) to filter 90% of noise with sub-200ms latency.

Ask & See – Retrieval-Augmented Document Intelligence System [GitHub Link](#)

- Developed a production-ready RAG application using FastAPI, LangChain, and ChromaDB for large documents.
- Deployed containerized microservices on Google Cloud Run with environment-based configuration management.

PUBLICATIONS

- Ocular Disease Recognition with Ensemble Techniques | ICCUBE, IEEE Xplore 2023 [Link](#)
- User Survey of Music Streaming Apps | HWWE, Springer 2023 [Link](#)