

Saharsh Sandeep Barve

 saharshbarve
 saharsh1005.github.io
 saharsh1005

 saharshbarve@gmail.com
 +1(217)-933-0383

EDUCATION

- **University of Illinois Urbana-Champaign**
• *MS in Computer Science; GPA: 3.96/4*

Urbana-Champaign, Illinois
Aug'23 - May'25

Relevant Coursework: Computer Vision, Cloud Computing, Applied Machine Learning, ML Algos for LLMs, 3D Vision, Adv. Data Management, Web Programming, Software Engineering

- **Manipal Institute of Technology**

Manipal, India

- *B.Tech in Computer Science and Engineering; GPA: 9.25/10*

Jul'18 - May'22

Relevant Coursework: Operating Systems, Parallel Programming, Computer Networks, Database Systems, Data Structure and Algorithms

EXPERIENCE

Dragonfruit AI

Software Engineer - ML Platform (Full-time)

Remote, USA / Menlo Park, CA
Jun'25 - Present

- **End-to-End ML Pipeline:** Led development and deployment of YOLO-based vision models, owning data pipelines, fine-tuning, evaluation, and customer-facing production releases.
- **Truck Turnaround Monitoring System:** Built a scalable computer vision pipeline for automated truck turnaround time monitoring using YOLO detection and license plate recognition; achieved ~300ms on-prem inference with cloud-integrated thumbnail storage.
- **Checkout Loss Detection:** Architected multi-signal checkout loss detection system integrating vision and transaction data; reduced false positives by 30% and accelerated review workflows.
- **Vector Deduplication & Optimization:** Implemented multi-threaded deduplication in Qdrant, cutting storage redundancy by 10%, and optimized pipelines for 20% higher throughput.

Onward Assist

Machine Learning Scientist (Full-time)

Bengaluru, India
Jul'22 - Jul'23

- **Nottingham Grading Tool:** Led the development of machine learning models such as Nottingham Scoring algorithm boosting breast cancer diagnosis accuracy by 30% compared to our previous baseline.
- **Deployment:** Integrated ML models into the web platform, utilized AWS and Kubernetes for scalable deployment, and optimized data workflows with MLflow and Apache Parquet.
- **HuBMap + HPA:** Led team to top 8% in HuBMap + HPA Kaggle competition, showcasing expertise in large-scale biomedical data analysis.

PROJECTS

- **Dynamic Prompting for LLMs:** Designed a pipeline for automated prompt selection in LLMs, optimizing efficiency via dynamic strategy selection (Zero-shot, Few-shot, CoT, SC-CoT). Achieved 89% accuracy with 5x fewer tokens and 5x faster runtime than SC-CoT. [\[github\]](#) [\[report\]](#)
- **Visual Odometry (Autonomous Vehicle):** Evaluated classical stereo vision and deep learning-based methods for visual odometry on the KITTI dataset, analyzing their efficacy in calculating depth maps and tracking motion. [\[github\]](#) [\[report\]](#)
- **3D Vision - Scene Reconstruction:** Conducted a study on 3D scene reconstruction concepts like Structure from Motion (SfM), Multi-View Stereo (MVS), and Neural Radiance Fields (NeRF). Explored Neural Kernel Surface Reconstruction (NKS) to refine NeRF results, tackling noise sensitivity. [\[report\]](#)
- **Graduate Researcher - UIUC (Virtual Reality, Computer Vision):** Worked on a medical instrument tracking system for HoloLens2, offering medical professionals real-time mixed reality guidance. (Aug'23 - Dec'23)

PUBLICATIONS

- **MSCS Thesis (arXiv):** Can we Debias Social Stereotypes in AI-Generated Images? Examining Text-to-Image Outputs and User Perceptions. [\[Link\]](#)
- **Paper:** Reef-Insight: A Framework for Reef Habitat Mapping with Clustering Methods Using Remote Sensing. Information 2023, 14, 373. [\[Link\]](#)
- **arXiv:** Switched auxiliary loss for robust training of transformer models for histopathological image segmentation. [\[Link\]](#)

SKILLS SUMMARY

- **ML Infrastructure & Agentic Systems:** End-to-end ML pipelines, Model orchestration, Evaluation frameworks, MLflow, TensorBoard, Experiment tracking, Production model deployment, A/B testing
- **AI/ML Frameworks & Tools:** PyTorch, TensorFlow, CUDA, HuggingFace Transformers, LangChain, LlamaIndex, AutoML, Synthetic data generation
- **Computer Vision & 3D:** DETR, YOLO (detection, tracking, fine-tuning), OpenCV, Object detection pipelines, 3D Reconstruction (NeRF, SfM)
- **LLMs & Prompting:** RAG pipelines, Zero/Few-shot orchestration, Prompt optimization, Agentic workflows
- **Production ML & Deployment:** Docker, Kubernetes, AWS (S3, ECS, Lambda, CloudFormation, RDS), CI/CD pipelines, On-prem inference optimization, Model monitoring, Scalable system design
- **Vector Databases & Search:** qdrant, Pinecone, FAISS, Embedding optimization, Deduplication
- **Programming & Engineering:** Python, C/C++, JavaScript, TypeScript, React, Node.js, REST APIs, Multi-threaded programming
- **Data & Databases:** SQL, NoSQL, Data pipelines, ETL workflows

MISCELLANEOUS

- Teaching Assistant, UIUC — CS444: Deep Learning for Computer Vision (Spring '24); CS441: Applied Machine Learning (Fall '24); CS498: Cloud Computing Applications.
- Open Source Contributor, [BespokeLabsAI/Curator](#) — Synthetic data curation for post-training & data extraction.
- Recipient, J N Tata Endowment Scholarship for academic excellence.
- Head of Finance, IAESTE India LC Manipal — Led financial operations for a 40-member organization.
- Mentored interns and led the Internship Training Program, Onward Assist (Jan '23 cohort).