PUSHKARAJ BARADKAR

Los Angeles, CA | +1 (213) 275-9348 | pushkarajbaradkar1@gmail.com | LinkedIn | Github | Portfolio

EDUCATION

University of Southern California, Viterbi School of Engineering GPA: 3.73 / 4.00 **Jan 2025 - Dec 2026**
Master of Science in Computer Science Los Angeles, CA
Machine Learning, Analysis of Algorithms, Database Systems, Security Systems, Side-Channel Threats

University of Mumbai GPA: 9.13 / 10 **Jul 2020 - May 2024**
Bachelor of Engineering in Computer Engineering Mumbai, India
Data Structures, Analysis of Algorithms, Operating Systems, Computer Networks, Machine Learning, Natural Language Processing

EXPERIENCE

AI Engineer Intern – Search and Recommendation Systems | Tabhi, Austin, TX **May 2026 – Aug 2026**
- Architected an end-to-end personalized discovery feed for 450K+ items using hybrid retrieval and slot-based re-ranking.
- Designed recency-decayed personalization over Elasticsearch BM25 and pgvector retrieval using behavioral signals.
- Reduced cold-feed latency from 8 s to 500 ms by three-tier caching, cache pre-warming, optimized indexing, and geospatial retrieval.
- Resolved Elasticsearch-MongoDB consistency issues and hardened location-aware retrieval, improving recommendation reliability.

Software Engineer | Leaniar LLC, Irvine, CA **Oct 2026 – Present**
- Building an LLM testing agent that generates, runs and repairs end-to-end tests from customer requirements, using Python and Django.
- Contributing to multi-tenant orchestration that runs 50+ agent workers in parallel over Redis pub/sub, with retries and audit logging.
- Working on an AWS Lambda and SQS pipeline that moves long-running jobs off the request path and streams results over WebSockets.

Robotics Software Engineer – Real-Time AI Systems | USC, Los Angeles, CA **Nov 2025 – Present**
- First-authored a paper (under review) and implemented a monocular SLAM pipeline for assistive glasses, using a 16-zone ToF sensor.
- Built GTSAM pose-graph backend in C++, reducing trajectory error from 29-58 m to 0.09-0.24 m while optimizing 2,486 poses in 0.36 s.
- Designed a physics-based ToF simulator and optimized zone pooling, halving trajectory error with 2-4% scale accuracy.
- Extending the factor graph to fuse IMU, ultrasonic, and GPS measurements for deployment on a Jetson Orin Nano.

Machine Learning Intern – Edge Computer Vision | YMT Medical, India **Jan 2024 – Aug 2024**
- Redesigned YOLOv8 with a P2 head and NWD loss for micro-lesion detection across 2,000+ patient images, achieving 0.83 mAP@0.5.
- Built a Laplacian/HSV image-quality pipeline to detect blur and glare in patient uploads before model inference.
- Optimized edge inference using INT8 quantization and operator fusion, reducing latency by approximately 40%.
- Mapped 468 MediaPipe facial landmarks to compute deterministic GAGS acne scores offline with an active learning loop.

Software Engineer Intern – Backend & APIs | Technoriya ERP Solution, India **Oct 2023 – Dec 2023**
- Developed Python and Node.js backend services and APIs supporting scalable data pipelines and cloud-based ML workflows.
- Implemented secure authentication and access protocols via Firebase Auth and Firestore rules to ensure data integrity.
- Optimized RESTful APIs and Firestore queries to achieve 450 ms response times under 1,000 concurrent requests for ML integration.

PROJECTS

Differentially Private Analysis of the UCI Adult Dataset | Python, diffprivlib, scikit-learn **Sept 2026**
- Applied three DP mechanisms to 30,162 census records at $\varepsilon = 1$ each, with budget accounting that blocks queries past $\varepsilon = 3$.
- Swept ε from 0.01 to 5.0: at $\varepsilon = 1$, counts were off by under 0.01% and DP logistic regression matched the non-private baseline at 0.784.
- Prevented two privacy leaks using a fixed public category list and schema-derived norm bounds independent of private data.

Warehouse Tote Perception – CV Pipeline | Python, FastSAM, OpenCV **Aug 2026**
- Built a four-stage FastSAM-x pipeline that locates tote items and picks a suction point, 0.958 pick success on 120 images vs. 0.31 baseline.
- Filtered 97 SAM masks/image to 6% valid items using a 2-of-3 color/edge vote and set cover (+0.21 F1, +0.27 pick).
- Cut inference from 20.6 s to 619 ms median on CPU/MPS by replacing MobileSAM with single-pass FastSAM.

HVAC Follow-Up Policy Optimization | Python, scikit-learn, offline RL **Aug 2026**
- Trained Fitted Q-Iteration on 74.7K logged steps for follow-up decisions, improving reward from -13.46 to +15.25.
- Distilled it into a depth-6 decision tree (77.8% agreement), shipping as one dependency-free file with escalation & message-cap guardrails.
- Evaluated 7 policies on 5K episodes via SNIPS, FQE, and doubly robust estimation, achieving +35.93 vs +8.48 reward.

TECHNICAL SKILLS, PUBLICATIONS, CONFERENCES

- **Programming Languages:** Python, C++, JavaScript, TypeScript
- **AI / ML:** PyTorch, Transformers, RAG, Agentic AI, Computer Vision, LangChain, LangGraph, MCP, Pandas, NumPy, scikit-learn
- **Frameworks & Development:** FastAPI, Flask, Node.js, Express.js, React.js, Next.js, REST APIs, OAuth, Git
- **Databases & Cloud:** PostgreSQL, pgvector, Elasticsearch, MongoDB, Redis, SQL, AWS, GCP, Azure, Docker, Kubernetes, CI/CD, Linux
- **Publications, Conferences, Hackathons:** P. Baradkar, L. Itti, SLAM research paper, under review; P. Baradkar et al., “Blockchain Based Music Streaming Platform Using NFTs,” IEEE ICSCSS 2023 [Cited by 7], ICSCSS ’23, SWE’25, SCBC’26 Finalist, Syrus ’23 Winner, Indehub ’25 Finalist