

Seshadri (Sesh) Sridharan

US Citizen · Greater Seattle Area · mit.seshadri@gmail.com · seshadri.xyz · linkedin.com/in/seshsridharan

Applied AI software engineer with 10+ years taking ML from prototype to production at Google, AKASA, and Walmart — shipping systems that reach hundreds of millions of users and automate millions of healthcare decisions. Founded and led ML/AI teams across search, GenAI, and automation.

EXPERIENCE

AI Founder @ [AlLever.io](https://aliover.io) & Foundations · Seattle

Oct 2024 – Jan 2026

Independently built and validated AI products through customer discovery and working prototypes — ranging from enterprise decision-making to consumer health.

- Group consensus agent: Map/reduce LLM architecture for structured stakeholder interviews at scale — aggregates and synthesizes recommendations for enterprise decision-makers.
- AI telephone agent for home service businesses — Inbound triage, Q&A, and booking for home service businesses with structured knowledge extraction and configurable escalation rules.
- Satellite + VLM address fingerprinting: GPS imagery + vision-language models for property-level object detection and demographic segmentation; validated with 6 enterprise customers.
- iOS nutrition tracking app: ARKit volumetric measurement + regression-based weight estimation + LLM nutrition extraction; deployed to live users.

Senior Software Engineer, TL @ Google · Health AI Summaries on Search · Seattle

Jul 2022 – Sep 2024

Technical lead for Google's first generative AI launch on Search for Health vertical (15% global Search traffic). Recognized with Spot Bonus award for org-level impact.

- Delivered hybrid RAG summarization system combining Knowledge Graph structured data with unstructured web snippets that improved relevance and user feedback metrics.
- Developed end-to-end triggering evaluation framework using LLM-as-judge — adopted as recommended standard across 10+ Search verticals for launch.
- Implemented multi-stage safety filtering (BERT classifiers → calibrated heuristics → fine-tuned LLM classifiers → rule-based guardrails) for medical content validation.
- Established AI governance process with medical doctors mapping clinical risk categories and safety review workflows; aligned with Google Trust & Safety frameworks.
- Led multi-quarter GenAI roadmap across engineering, product, research, and safety orgs; partnered with Google Research to adapt MedPaLM for Search.
- Founded weekly GenAI knowledge-sharing series upskilling 30+ engineers across Search verticals; consulted as Applied ML Advisor to 6+ teams.

Lead Machine Learning Engineer @ AKASA · Healthcare AI · San Francisco

Jan 2020 – Jul 2021

Founding ML engineer. Built foundational MLOps platform, human-in-the-loop architecture, and research collaboration from scratch at a healthcare revenue cycle automation startup.

- Developed a universal field extraction system using model ensembles + active learning — eliminated site-specific parsers, reduced labeling costs by 70% and enabled new clients to go live in days vs. weeks.
- Designed human-in-the-loop architecture with configurable per-tenant automation thresholds — achieved 80+% task completion at 97+% accuracy across millions of monthly claims.
- Delivered multi-tenant ML training pipeline, shadow deployment + human in the loop active eval framework, statistical drift detection, and automated retraining triggers.
- Led research collaboration published at ICML 2020 (DeepClaim: Payer Response Prediction); drove hiring, cloud cost optimization, and MLOps best practices.

Senior Software Engineer, TL @ Google · Search & Brain · Mountain View

Nov 2015 – Jan 2020

Technical Citizenship award for org level impact. Led hardware accelerator integration on Brain stealth team and founded a new ML product team within Search.

- Founded Travel Experiences team: Won VP-level approval, hired 5 engineers, and overhauled travel entity indexing from heuristics to production ML pipeline (using geometric, location history, web, engagement features) — resulted in a landmark XX% lift to recall, F1-score and user engagement.
- Developed hybrid human-ML travel experiences system; organized cross-office hackathon (35 engineers, 10 teams) that shaped the roadmap for 6 subsequent quarters.
- Brain stealth team: Led client integration for new hardware accelerator — delivered inference simulator framework adopted by Translate, Image Search, and Assistant to validate accelerator performance against real production workloads.

Senior Software Engineer @ Walmart Labs · Ecommerce Search · Sunnyvale

Jun 2013 – Oct 2015

Co-led relevance and multi-tenancy architecture for ecommerce search platform serving walmart.com, samsclub.com, and grocery.walmart.com.

- Delivered multi-tenancy infrastructure consolidating three retail properties onto a unified serving layer with cross-property signal transfer for relevance improvement, eliminating cold start.
- Developed search space pruning algorithms reducing query execution time significantly, saving millions in compute costs.
- Built learning-to-rank, item similarity and query understanding signal pipelines — delivered durable revenue and user engagement lift validated via A/B testing.

SELECT TECHNOLOGIES

AI / ML	scikit-learn, MLflow, DVC, Prefect, PyTorch, TensorFlow, LangGraph, Agentic AI (ReAct, MCP, RAG), BAML
Infra	GCP (Vertex AI, BigQuery), AWS (SageMaker, ECS), Kubernetes, Docker, Postgres, DataFlow, Grafana
Languages	Python, C++, Java

EDUCATION

M.S. Language Technology (AI), Computer Science — Carnegie Mellon University	2013
B.S. Computer Science — Manipal University, India	2011

PATENTS

- Systems and methods for understanding search queries (US20170221118A1)
- System, method, and non-transitory computer readable media for generating synonyms of a search query (US20190324979A1)

PUBLICATIONS

- Deep Claim: Payer Response Prediction from Claims Data with Deep Learning | ICML HSYS Workshop 2020
- Conversational Strategies for Robustly Managing Dialog in Public Spaces | EACL 2014
- Situated Multiparty Interaction Between Humans and Agents | ACM ICMI 2013
- Modeling Word Meaning: Distributional Semantics and the Corpus Quality-Quantity Trade-off | COLING CogALex 2012
- NeuroDialog: An EEG-Enabled Spoken Dialog Interface | ACM ICMI 2012

HONORS

- Technical Citizenship Award, Google Travel & Shopping Organization
- Multiple Spot Bonus Awards, Google: AI Summaries launch · Travel Indexing · Experiences · Hackathon
- Most Scalable Hack, Code For India Hackathon 2015
- Graduate Research Fellowship, Carnegie Mellon University