SAIRAM KRISHNAN

Seattle, WA | Open to relocation | U.S. Citizen | 281-513-1263 | sairambkrishnan@gmail.com
sairambkrishnan.com | linkedin.com/in/sairambkrishnan | github.com/SaiWebApps

PROFESSIONAL EXPERIENCE

Apple | Principal Software Engineer — Feb 2021 - Present
Apple Media Products | Team Lead since 2022 | Seattle, WA

- Manage eight engineers, five of them senior, and seven vendors. Own performance reviews and conversations, ratings and calibration input, and compensation recommendations; lead hiring and mentor engineers toward promotion.
- Architected and delivered GraphRAG evaluation and correction for Gemini and internal foundation models across music, books, and podcasts: graph-grounded fact checking, structured error reports and diagnostic hypotheses fed into subsequent queries, and separate checks for graph contradictions.
- Raised data quality 40% by leading supervised ML entity resolution. Defined candidate generation, features, labels, evaluation, and human overrides with ML engineers; integrated classifiers, corrected artist identities, and transferred ongoing ownership.
- Directed a 15-engineer program across four offices to launch Apple Music Credits; built graph infrastructure for Apple Music Classical. Unified 15+ sources in the media graph powering search, recommendations, and discovery for 100M+ users.
- Cut average processing time for millions of tracks from approximately 60 hours to one minute by leading the migration from Spark to Flink, Kafka, and Cassandra. Won Ops, PM, and client backing to launch a credits feature on the new platform, consolidating batch, streaming, and instant-curation logic for partial and complete edits.
- Built Crucible, an agent engineering framework for operational diagnosis and end-to-end testing, with Model Context Protocol tools, multiagent review, model-based action checks, and executable acceptance rules. QA, Ops, and partner teams extended it into cross-service tracing and end-to-end tests.
- Direct senior and junior engineers delivering ML-generated media tags and awards metadata; designed versioned predictions, persistent human overrides, and client integrations to preserve editorial decisions through model updates.
- Coached a senior engineer into independent load-testing and performance ownership across 1M–100M-message workloads through prototype-driven decisions and incremental delivery.
- Increased velocity 30% and reduced incidents 45% by establishing design-review and CI/CD standards.

Amazon | Senior Software Engineer — Apr 2018 - Feb 2021
Alexa Core Engine | Promoted from L5 to L6 | Seattle, WA

- Managed eight engineers; owned performance reviews, hiring, 1:1s, and promotion recommendations. Mentored half the team to promotion; selected for Amazon's Mentorship and Leadership Development Program.
- Improved end-to-end automatic speech recognition (ASR) and skill-routing accuracy by 40% through skill-aware prefetching and processing near ASR, feeding linked compound-intent context back into recognition and hypothesis selection.
- Designed a GraphQL aggregation and orchestration platform saving $40M annually by eliminating redundant processing through annotation-driven rules, batched results, prefetching, and skill-specific cache keys.
- Cut onboarding from 2 weeks to 15 minutes via self-service deployment; reduced weekly operational issues from 40 to 10.
- Secured fleet investment and directed Prime Day and Black Friday capacity forecasting. Redesigned new-user registration, reducing peak-holiday registration issues tenfold.

Susquehanna International Group | Senior Financial Data Engineer — Nov 2016 - Mar 2018
Promoted from Financial Data Engineer | Philadelphia, PA

- Led real-time FINRA compliance surveillance across market-making transactions in C++, Python/Pandas, and Django, flagging violations at execution.
- Automated capacity scaling with real-time hardware asset tracking, resolving chronic inter-team coordination failures.

Microsoft | Software Development Engineer — Mar 2015 - Oct 2016
Windows OS Group | Redmond, WA

- Cut Windows 10 Mobile Mail power consumption 4× versus Windows Phone 8; set the app's power targets.
- Used real-time telemetry dashboards to brief leadership weekly and choose the default email auto-sync interval.

EXECUTIVE EDUCATION & ADVISORY

Praxtera AI Institute | Head of AI & Executive Education — Sep 2025 - Present
Curriculum, advisory, and speaking | Concurrent with Apple

- Own curriculum and serve as principal AI subject-matter expert and speaker for HP, Morgan Lewis, PBS, and Sonova; lead client discovery and AI solution architecture.
- Enabled HP teams to automate recurring work with Copilot workflows; taught prompting and workflows at Sonova and prompting fundamentals to Jacobson Strategic's team.
- Trained Morgan Lewis's internal AI instructors, using AI-assisted personas and live roleplay to address adoption resistance.
- Led a PBS session on AI strategy with privacy and governance first, and on automating work that takes staff time away from the mission.
- Keynoted "AI for Impact" for the Philadelphia Eagles at the Eagles Care Summit, Lincoln Financial Field (March 18, 2026), reaching 300+ community and business leaders.

INDEPENDENT SPEAKING & TEACHING

Atos | "Knowledge Graphs, Then Small Language Models" Sep 3, 2026

- Presented how graphs enrich questions and check generated answers, how to choose model size, and how to set up a local model live.

Carnegie Mellon University | Web Applications Course Developer May 2013 - Dec 2013

Teaching Assistant

- Redesigned the course with the professor, planning a 20-lecture semester and building client-side and server-side exercises and reference applications in Spring MVC, Play, and Django. Instructed 105+ students as Agile Product Owner for their final projects.

AI INVESTMENT & WORKFORCE RESEARCH

A Multi-Layer Framework for Evaluating the Return on Investment of AI Projects

Sairam Krishnan, Sonia Gandhi, Andrea Hepp. SSRN working paper, 2026. doi:10.2139/ssrn.6732598

Integrates readiness, ROIC value drivers, staged investment, and a decision scorecard with eight categories of AI risk, including vendor dependency and organizational friction. Eight illustrative public-company cases examine inference costs and organizational readiness. Wharton independent study; faculty sponsor Professor Gad Allon.

Who Trains the Next Senior? Managing Software Engineering When AI Makes Code Cheap

Gajendra Babu Thokala, Sairam Krishnan. SSRN working paper, 2026. doi:10.2139/ssrn.7083078

Examines "review theater": approvals that no longer meaningfully verify AI-generated work. Proposes smaller reviewable changes, scrutiny matched to consequence, deliberate apprenticeship, and measurement of verified outcomes rather than code volume.

HUMAN-SUPERVISED SYSTEMS RESEARCH

Drone-Megaddon 2026: Co-Intelligent Swarm Control with Language and Vision Models

Sairam Krishnan, Daniel Ting. engrXiv preprint, 2026. doi:10.31224/6987

Proposes language-intent, visual-grounding, and multi-drone coordination layers with operator inspection gates, audit logs, and manual-control fallback under degraded connectivity. Illustrated through a search-and-rescue scenario.

Drone-Megaddon: A Real-Time Strategy Interface for Managing Drone Swarms

Sairam Krishnan, Daniel Ting. Carnegie Mellon course project, 2014; archived on engrXiv, 2026. doi:10.31224/6924

Built an end-to-end Android, Zigbee radio, and embedded-control prototype using a real-time strategy interface.

TECHNICAL SKILLS

Java, Scala, Python, Go, C++, TypeScript; Spark, Flink, Kafka, Cassandra, Neo4j, GraphQL, Protobuf, AWS; supervised ML pipelines, GraphRAG, LLM evaluation, MCP, agent orchestration.

VENTURES & INDEPENDENT AI PROJECTS

Ondoway | Co-founder & Chief Technology Officer

AI travel companion | Private beta | Concurrent with Apple

- Own product strategy and architecture: graph-grounded storytelling, deterministic routing, and location-triggered narration using Flutter, FastAPI, and Neo4j. Turned beta feedback in New York and Paris into route-sequencing priorities.
- Established LLM evaluations with frozen human-labeled cases, blinded comparisons, evidence checks, and code-enforced gates; rejected a cheaper LLM judge after calibration exposed false rejections of valid paraphrases.
- Instrumented model/prompt identity, tokens, latency, responses, and evidence; tracked generation and narration costs including failed attempts to guide debugging, model selection, and unit economics.

Persuaider | Creator & AI Engineering Lead

AI negotiation simulation and coaching | Independent work

- Created interactive practice and autonomous multiagent negotiation with distinct motivations, confidential objectives, and walk-away limits.
- Extracted negotiated terms with LLMs; scored utility, deal feasibility, and outcomes deterministically against fixed rules.
- Designed configurable coaching rubrics; checked feedback against transcript evidence, retried failures, and withheld unsupported coaching.

Gestiq | Creator

On-device computer vision | Independent work

- Developed a native macOS gesture-control application extending a CMU capstone: a hand-controlled 2048 puzzle game.
- Implemented temporal recognition over Apple Vision hand poses, normalizing motion to palm size for two-hand controls and context-aware commands.
- Discarded stale frames, rechecked active targets, and added live diagnostics to guard desktop actions against uncertain camera input; video stays on device.

EDUCATION & HONORS

The Wharton School, University of Pennsylvania | Executive MBA May 2026

Palmer Scholar (top 5%); **Benjamin Franklin Award** (peer-voted, one of three recipients in a class of 250); Director's List in all six terms; First Year Honors. Five majors: AI in Business; Strategic Management; Operations, Information & Decision-Making; Entrepreneurship & Innovation; Business Analytics.

Carnegie Mellon University | M.S. & B.S., Electrical & Computer Engineering 2014

M.S., Dec 2014 | B.S., May 2014

M.S. GPA: **4.0/4.0**. B.S. minor: Business Administration. Dean's List, Carnegie Institute of Technology, 2012-2014.