# SAIRAM KRISHNAN

Engineering Leader · Scaling a 100M-User AI Knowledge Graph · Wharton MBA
Seattle, WA · Open to relocation · 281-513-1263 · sairambkrishnan@gmail.com · U.S. Citizen
linkedin.com/in/sairambkrishnan · github.com/SaiWebApps · orcid.org/0009-0004-2626-1273

## PROFILE

Engineering leader with 11 years of experience, rising from junior software engineer at Microsoft to senior software engineer at SIG and ultimately, engineering lead in Amazon and Apple. Currently a Principal Software Engineer at Apple, leading an 8-engineer team and scaling a media knowledge graph for 100M+ users. Personally ships production agentic AI, including an LLM ops assistant the team uses daily. Earlier at Amazon, led 8 engineers and built the Alexa Core Engine, a platform that saved $40M a year and lifted model accuracy 40%. Also achieved Wharton MBA in May 2026 and teaches AI strategy to C-suite leaders on the side through Praxtera AI Institute.

**Targeting a Senior Software Engineering Manager role in AI/ML.**

## EXPERIENCE

**Apple** | Principal Engineer & Senior Engineering Leader, Knowledge Graph Platform — Feb 2021 – Present

*Apple Media Products · Seattle, WA · Engineering lead for an eight-engineer team across Seattle, London, and partner sites in Europe and Latin America, reporting to the engineering director*

- Led architecture and cross-functional delivery of Apple's media knowledge graph, the canonical model unifying 15+ sources to power search, recommendations, and discovery for over 100 million users.
- Mentored engineers to promotion, shaped performance reviews, and drove hiring; set the design-review and CI/CD standards that lifted team velocity 30% and cut production incidents 45%.
- Built and shipped production agentic AI: an LLM ops assistant for debugging logs, traces, and metrics over MCP in natural language, and a skill platform that turned pipeline work into plain-English workflows for daily, on-call use.
- Led the four-office global launch of Apple Music Credits (Cupertino, Seattle, London, Singapore), coordinating a 15-engineer cross-functional program to deliver real-time credits curation for millions of tracks.
- Drove the platform from batch to real-time streaming, re-architecting ingestion, similarity mapping, and graph generation from Spark onto an event-driven Flink, Kafka, and Cassandra stack.
- Designed the ML entity-resolution system pairing similarity mapping with clustering to deduplicate entities across sources, lifting data quality 40% while cutting downstream cost.
- Owned the graph query services and the client library for partner teams to build on, and set the shared Protobuf contracts adopted across the platform.

**Praxtera AI Institute** | CTO and Curriculum Designer, AI/ML Executive Education — Sep 2025 – Present

*Philadelphia-based AI executive-education firm · concurrent with the Apple role*

- Taught AI strategy to C-suite leaders at HP, Morgan Lewis, PBS, and the Philadelphia Eagles, working from an executive curriculum built on original Wharton research to help executives evaluate, govern, and deploy AI.
- Keynoted the Philadelphia Eagles' 12th annual Eagles Care Summit at Lincoln Financial Field in 2026, delivering "AI for Impact" to more than 300 community and business leaders.

**Amazon** | Senior Software Engineer, Alexa — Apr 2018 – Feb 2021

*Seattle, WA · Promoted from SDE II (L5) to Senior SDE (L6); led an eight-engineer team on the Alexa Core Engine*

- Designed a GraphQL data-aggregation and caching platform for the Alexa Engine whose annotation-driven rule engine batched and clustered results to the right downstream services, saving $40M a year and lifting model accuracy 40%.
- Mentored half the team from entry level to promotion, earning selection for Amazon's Mentorship and Leadership Development Program.
- Collapsed partner onboarding from two weeks to 15 minutes with a self-service rapid-deployment platform.
- Owned operational excellence, driving the weekly issue count from 40 to 10 and re-engineering the new-user registration flow to cut peak-holiday customer issues tenfold.
- Led capacity planning for Prime Day and Black Friday, forecasting hardware demand and winning leadership's commitment to expand the fleet ahead of peak.

**Susquehanna International Group (SIG)** | Senior Financial Data Engineer Nov 2016 – Mar 2018

*Philadelphia, PA · Promoted from Financial Data Engineer*

- Led real-time FINRA compliance surveillance across the firm's market-making transactions in C++, Python/Pandas, and Django, flagging violations at the moment of execution.
- Designed and shipped a real-time hardware asset-tracking system that automated capacity scaling and ended chronic inter-team coordination failures.

**Microsoft** | Software Development Engineer, Windows OS Group Mar 2015 – Oct 2016

*Redmond, WA*

- Cut Windows 10 Mobile Mail power consumption 4x against the Windows Phone 8 baseline, and set the app's power targets.
- Drove product decisions with telemetry: real-time dashboards, weekly guidance to leadership, and the data-derived default for the auto-sync interval.

## EDUCATION

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

Palmer Scholar (top 5% of class) · Benjamin Franklin Award (peer-voted) · Director's List, all six terms · First Year Honors

Majors (5): AI in Business · Strategic Management · Operations, Information & Decision-Making · Entrepreneurship & Innovation · Business Analytics

**Carnegie Mellon University** | M.S. in Electrical & Computer Engineering, GPA 4.0/4.0 Dec 2014

**Carnegie Mellon University** | B.S. in Electrical & Computer Engineering; Minor in Business Administration May 2014

## PUBLICATIONS

- Krishnan, S., & Thokala, G. B. (2026). *Who Trains the Next Senior? Managing Engineering Teams When AI Makes Code Cheap*. Working paper.
- Krishnan, S., Hepp, A., & Gandhi, S. (2026). *A Multi-Layer Framework for Evaluating the Return on Investment of AI Projects*. SSRN. doi:10.2139/ssrn.6732598
- Krishnan, S., & Ting, D. (2026). *Drone-Megaddon 2026: Co-Intelligent Swarm Control with Language and Vision Models*. engrXiv. doi:10.31224/6987
- Krishnan, S., & Ting, D. (2014). *Drone-Megaddon 2014: A Real-Time Strategy Interface for Managing Drone Swarms*. engrXiv (archived 2026). doi:10.31224/6924

## INDEPENDENT R&D

*Self-directed R&D: production-grade systems built and shipped solo, from architecture to deployment.*

- **DependIQ:** The verification layer agentic coding still lacks. Maps the blast radius of any dependency change across a repo graph, then generates targeted, reviewable fixes. Built on FastAPI, Neo4j, and Postgres.
- **Crucible:** Adversarial QA for AI agents, open-sourced. Rival agents attack a problem from opposing angles behind an evidence-gated review, and hooks kill unsafe tool calls before execution. A Claude Code toolkit.
- **Persuaider:** A sparring partner for high-stakes negotiations. In-character personas push back in real conversation, then an LLM scores the full transcript and surfaces the moves that land. Built on Next.js, with Claude, GPT, and Gemini behind a failover chain.
- **Ondoway:** Personalized audio tours, composed on the fly. A Neo4j knowledge graph (7 entity, 11 relationship types) models travelers, places, and stories; an LLM and text-to-speech narrate each trip as travelers move. FastAPI backend, 177 tests.

## SKILLS

**Leadership:** Engineering team leadership · Cross-functional and multi-site delivery · Mentorship and talent development · Hiring · Roadmap ownership · Executive communication and public speaking

**Applied & Agentic AI:** LLM application development (Claude API) · Agentic systems and MCP · Multi-provider LLM orchestration · AI-assisted engineering operations · AI strategy and ROI evaluation

**Data & ML Platforms:** Knowledge graphs · Entity resolution · Flink · Kafka · Spark · Cassandra · Neo4j · GraphQL · Protobuf/Avro · AWS · Distributed systems at 100M-user scale

**Programming Languages:** Java · Scala · Python · Go · C++ · TypeScript/JavaScript

**Spoken Languages:** English · Tamil · Spanish · French · Italian · Chinese