

# RAM KADIYALA

☎ +91 8919628977 | ✉ [contact@rkadiyala.com](mailto:contact@rkadiyala.com) | 🌐 [1024-m.github.io](https://1024-m.github.io) | 🐦 [Twitter/X](#) | 📄 [GitHub](#)

## SUMMARY

---

Founding Solution Architect hire for agentic systems at two startups across the US and India. Early agent expertise from before public release and experience across engineering, research, product, marketing and sales.

Seeking Solution Architect / Tech Lead roles focused on Gen AI and agents.

## EDUCATION

---

M.S., Machine Learning,	University of Maryland, College Park, US	2022 - 2024
B.Tech., Computer Science,	RVR & JC College of Engineering, IN	2018 - 2022

## SKILLS

---

AI / ML	PyTorch, Transformers, Hugging Face, fine-tuning (LoRA, SFT, DPO, GRPO, RLHF, ..), classical ML, Time series, NumPy, Pandas
Agent Frameworks	Strands, Google ADK, MCP, A2A, LangChain, LangGraph
RAG & Memory	Semantic search, Multimodal embeddings, FAISS, Memory Compression, Re-ranking
Cloud / LLMOps	AWS, GCP, Docker, WandB, Guardrails, Observability, Governance, Compliance
Web & Frontend	Three.js, FastAPI, Gradio, Streamlit, React.js, Node.js
Client-facing & Delivery	solution design, system design, technical pre-sales, stakeholder management, requirements gathering, proof-of-concepts, PRDs & SoWs, delivery management.

## EXPERIENCE

---

AI Solution Architect Aivar	Nov 2025 – Present <i>Bangalore, IN</i>
--------------------------------	--

Leading solution design, building end-to-end production systems, and delivery of agentic and LLM-based systems for high-value enterprise clients on AWS across India, US and Middle East.

- **PineLabs – Semi-autonomous SRE Agent:** Self-healing incident resolution across L1–L3 tiers with HITL feedback loops and engineer annotation-driven continual improvement.
- **HighRadius – Infra Resource Allocation Agent:** Autonomous provisioning and right-sizing of cloud resources with demand-driven scaling and engineer-in-the-loop override controls.
- **AGS Health – Healthcare Claims Appeal Agent:** End-to-end automation of claims appeal workflows processing up to 30 document types and 300+ page inputs/outputs via multi-agent pipelines.
- **Swiggy – Membership Conversion Forecasting:** SageMaker-based pipeline with automated A/B testing and best-model selection to predict user membership purchase likelihood.
- **M2P Fintech – Banking File Parser:** One agentic pipeline to parse and standardize payment and banking documents across thousands of formats from multiple financial institutions across the globe.

Open Science Community Lead Cohere Labs	May 2025 – Present <i>Remote</i>
--	-------------------------------------

Lead the open-source research program across two groups – Community Projects and ML-Agents – covering LLM post-training and agentic systems at production scale.

- **Program Leadership:** Leading the open-science program across both groups, finding research gaps, setting direction, and coordinating distributed contributors. [labscommunity.cohere.com/community-leads](https://labscommunity.cohere.com/community-leads)
- **Grants & Funding:** Secured over 50K+ USD in compute and funding grants from Anyscale, Hugging Face, and Modal sustaining community project requirements at production scale.
- **Pre-Launch & GTM:** Beta-testing feedback and go-to-market input on new Cohere model releases, bridging the labs community and product teams.

- **Mentorship:** Mentoring undergraduates and early-career researchers on their research ideas, with 12+ papers published through the program at tier-1 venues: ACL, EMNLP, NeurIPS, etc.

**NLP Research Lead**  
**Traversaal.ai**

**Sep 2023 – Aug 2025**  
**San Francisco, US**

Delivered production agentic systems and AI products for Fortune 500 clients while leading NLP research on multilingual models and agent evaluation benchmarks.

- **Home Depot – Route Optimization Agent:** Built an agent for delivery route planning that incorporates user-specified constraints such as time windows, evening-only deliveries, and blackout periods.
- **Fortune 500 Clients – Agentic OS:** Developed an organization-wide agentic operating system for data access, workflow automation, and orchestration; deployed in production for multiple Fortune 500 companies. [olive](#)
- **Enterprise Clients – Web Scraper & Search Engine (Ares):** Built a scalable web scraping and semantic search engine for enterprise data ingestion and retrieval; adopted by several large organizations. [ares](#)
- **Research Initiatives – Multilingual LLMs & Benchmarks:** Curated synthetic datasets for Indian languages with localized context; developed benchmarks and frameworks for fine-grained evaluation and continual learning of domain-specific agents.
- **Course Contributor:** Contributed to curriculum development for the “Building and Scaling AI Solutions” course, Spring 2025, Stanford Continuing Studies.

**Founder**  
**Black Ops Esports**

**Nov 2018 – Apr 2022**  
**Hyderabad, IN**

Founded and scaled a mobile-gaming esports organization spanning competitive teams, a 60K+ community, a monetized regional creator network, and data-driven predictive modeling for in-game strategy.

- **Community Growth:** Grew the organization’s social presence organically with \$0 ad spend to 56K on [Instagram](#) and an 18K member [Discord](#) community.
- **Creator Network & Monetization:** Built and managed a network of regional gaming content creators across multiple Indian languages – connecting them with advertisers, guiding content strategy and collaborations, and earning a share of the ad revenue generated.
- **Competitive Teams:** Recruited and managed competitive esports teams of 200+ members across several mobile gaming titles.
- **Game Analytics:** Built game-progression forecasting models for battle-royale titles to simulate match dynamics and stress-test drop and rotation strategies against projected placement outcomes.

## PROJECTS

---

1. **Indic Gemma 4.** Trained 23 regional-language LLMs (27B MoE) for Indian languages, built on a synthetic dataset of 23M samples and 2.3T tokens with localized cultural context. ([Collection](#))  
*Stack: PyTorch, Transformers, Hugging Face, DynamoDB, ...*
2. **Street Navigation Agent.** An agentic pipeline that uses a VLM to explore Street View imagery across a region and return evidence-backed findings for free-text place queries; built on a 8-tool orchestration with a shared Hugging Face dataset cache (race-safe, last-writer-wins) that cuts Street View API quota usage. ([Repo](#))  
*Stack: Strands, OpenRouter, Google Maps API, Hugging Face, ...*

## CERTIFICATIONS

---

**AWS Certified Machine Learning - Specialty (MLS-C01)**  
**NVIDIA Accelerated Computing with CUDA**

 [Certificate](#)  
 [Certificate](#)

Valid till Mar 2029  
Valid till Apr 2027

## PUBLICATIONS

[G](#) [GOOGLE SCHOLAR](#)

---

1. **Global PIQA: Evaluating Physical Commonsense Reasoning Across 100+ Languages and Cultures**  
*NeurIPS 2026* [Paper](#)
2. **Kaleidoscope: In-language Exams for Massively Multilingual Vision Evaluation**  
*ICLR 2026* [Paper](#)
3. **DSBC: Data Science task Benchmarking with Context engineering**  
*AAACL 2025 Main* [Paper](#)
4. **Robust and Fine-Grained Detection of AI Generated Texts**  
*TMLR 2025* [Paper](#)
5. **Improving Multilingual Capabilities with Cultural and Local Knowledge in Large Language Models While Enhancing Native Performance**  
*AAACL 2025 Main* [Paper](#)
6. **Uncovering Cultural Representation Disparities in Vision-Language Models**  
*AAACL 2025 Findings* [Paper](#)
7. **Evaluating Generalization Capabilities of LLM-Based Agents in Mixed-Motive Scenarios Using Concordia**  
*NeurIPS 2025 D&B* [Paper](#)
8. **Query Attribute Modeling: Improving search relevance with Semantic Search and Meta Data Filtering**  
*ICDM 2025* [Paper](#)
9. **Lexical Reranking of Semantic Retrieval (LeSeR) for Regulatory Question Answering**  
*COLING 2025* [Paper](#)
10. **Sequential Learning for Claim Verification and Explanation Generation in Financial Domains**  
*COLING 2025* [Paper](#)
11. **NLU of Devanagari Script Languages: Detection of Language, Hate Speech, and Targets using LLMs**  
*COLING 2025* [Paper](#)
12. **Augmenting Legal Decision Support Systems with LLM-based NLI for Analyzing Social Media Evidence**  
*EMNLP 2024* [Paper](#)
13. **Cross-lingual emotion detection through Large Language Models**  
*ACL 2024* [Paper](#)
14. **Self Reported Health Text Classification through Ensembles**  
*ACL 2024* [Paper](#)
15. **Black-Box Word-Level Text Boundary Detection in Partially Machine Generated Texts**  
*NAACL 2024* [Paper](#)
16. **Multi-class Emotion detection on highly imbalanced data**  
*ACL 2023* [Paper](#)

## CO-CURRICULAR

---

- **Program Committee Member** – Workshop on Detecting AI Generated Content @ COLING 2025.
- **Reviewer for Conferences** – NeurIPS 2024 to Present, ACL ARR: 2023 to Present.

## INVITED TALKS

---

1. **Ontario Tech University** *June 2026*  
Invited talk on ongoing research in Agentic AI, evals, and benchmarking.
2. **Cohere Labs Summer School** *August 2026*  
Invited talk on Agents 101, industry trends, potential research directions.