

Sai Koti Reddy Danda

AI Research and Engineering Leadership

Location: Bengaluru, India
Email: d.saikotireddy@gmail.com
LinkedIn: linkedin.com/in/saikotireddy
Website: saikotireddy.com
Scholar: Google Scholar profile

Profile

AI research and engineering leader with **16 years in AI research** and a record of building enterprise AI products from the ground up in two startups. I have led **multiple products and distributed teams across three geographies**, setting technical direction with founders, executives, and customers. Products I led are used by **Fortune 500 enterprises** including NVIDIA, JPMorgan Chase, CVS Health, Kroger, and T-Mobile, and banking products I led from inception are in proof-of-concept engagements with **SMBC**.

- **Product portfolio:** Originated and led AI products spanning banking process intelligence, agentic AI, Kubernetes autoscaling, GPU orchestration, resource optimization, and intelligent load balancing
- **Teams and organization:** Managed **three AI teams across India, Romania, and Canada**; recruited, onboarded, and mentored engineers; built research partnerships with IIT Bombay and IIIT Bangalore
- **Outcomes:** **More than 30% lower cloud costs** through reinforcement learning (RL) driven autoscaling, and **3x LLM inference throughput** through GPU orchestration
- **Research and IP:** Co-wrote **two funded five-year research grants** (IBM Research Math Council; Department of Science & Technology, India and Robert Bosch Centre, IISc); peer-reviewed publications at AAMAS, IEEE CDC, and ECAI; contributions to multiple U.S. patents and patent applications; IBM awards for research and client value

Research depth in multi-agent systems, agentic AI, large language models (LLMs), and reinforcement learning (including constrained and risk-averse multi-agent RL), with foundations in game theory and stochastic optimization.

Experience

kAIgentic Inc

Bengaluru, India

Principal Architect – Agentic AI for banking process transformation
Leadership and Scope

Dec 2025 – Present

- Lead AI research and architecture across kAIgentic’s banking product suite from early ideation; set technical direction across teams with product, technology, and customer success leaders, turning customer priorities into research and engineering decisions

Key Contributions

- Originated products for process mapping, process transformation, and agentic AI, translating enterprise banking problems into research priorities, prototypes, and product capabilities
- Architected multi-agent process intelligence that combines multiple information sources, banking ontologies, and knowledge graphs for process mapping and analysis

Impact

- Products I have led from inception are in proof-of-concept engagements with **SMBC** for banking process transformation

Avesha Inc

Bengaluru, India

Principal AI Architect – Employee #5, joined at seed stage
Leadership and Scope

Mar 2021 – Dec 2025

- Led AI research and product creation across a portfolio of **five products**, taking each from concept through proof of concept and production in enterprise environments

- Managed **three AI teams across India, Romania, and Canada**; recruited, onboarded, and mentored engineers; established the company's U.S. AI presence after its Series B
- Led all AI initiatives, working directly with the CEO and executive leadership on product strategy, customer acquisition, and technical direction
- Initiated and led university research collaborations with IIT Bombay and IIIT Bangalore

Impact

- RL-driven autoscaling (Smart Scaler) reduced cloud costs by **more than 30%**; GPU orchestration (Elastic GPU Service) increased LLM inference throughput **3x**
- Products used by **Fortune 500 enterprises** including NVIDIA, JPMorgan Chase, CVS Health, Kroger, T-Mobile, and Supermicro, and by customers across telecom, media, security, cloud infrastructure, and financial services including Cox, Akamai, Nutanix, Amagi, McAfee, and Finvi; technology partners included NVIDIA, Nutanix, and Oracle
- Products scaled across 100+ customers, 56+ clusters, and 10,000+ applications
- Contributed inventions in workload placement (granted U.S. patent), adaptive autoscaling, and RL-based global load balancing (patent filings)

Products

- **Smart Scaler** (RL-based Kubernetes autoscaling), **Elastic GPU Service** (multi-cloud GPU orchestration), **Avesha Obliq** (agentic SRE and QA automation), **SmartSizing** (container resource optimization), and **Smart Application Load Balancer** (RL-based service-mesh routing)

IBM Research India

Bengaluru, India

Research Engineer – AI for Supply Chains Group

Jul 2017 – Feb 2021

Key Contributions

- Core contributor to AI-based analytics for Sterling Commerce Multi-Enterprise Business Networks, automating decisions in multi-party supply chains
- Co-wrote the five-year IBM Research Math Council grant *Making Game Theory Practical for Decision-Making in Multi-Agent Systems*, and led game-theoretic research in constrained and risk-averse multi-agent reinforcement learning under it
- Applied reinforcement learning, game theory, and optimization to inventory pooling, lateral transshipment, and collaborative shipping
- Led development of the IPDC supply chain and dealer finance blockchain platform and contributed to blockchain solutions for Bajaj Auto and M&M Finance

Impact

- Improved invoicing efficiency and transparency for Bajaj Auto and streamlined financial processes for M&M Finance
- Math Council grant work led to publications on risk-averse and constrained multi-agent RL (AAMAS 2019, AAMAS 2021), inventory pooling (IEEE SCC 2022), and truthful lateral transshipment (ECAI 2023); further publications at IEEE Blockchain, and patent filings in AI and blockchain
- IBM Research Spotlight Award (2020), Client Value Outstanding Technical Achievement Award (2019), and Research Accomplishment Award (2018)

Robert Bosch Centre for Cyber-Physical Systems, IISc

Bengaluru, India

Research Assistant – Stochastic Systems Lab, Prof. Shalabh Bhatnagar

Dec 2016 – Jun 2017

- Co-wrote the five-year grant *Distributed Multi-Agent Algorithms for Dynamic Control of Microgrids*, funded by the Department of Science & Technology, India and the Robert Bosch Centre for Cyber-Physical Systems
- Researched distributed reinforcement learning for decentralized microgrid control; the grant supported publications on microgrid energy management (IEEE SmartGridComm 2018, IEEE ISGT-Europe 2020), stochastic gradient search (IEEE CDC 2018), and constrained multi-agent RL (AAMAS 2021)

IBM Research India

Research Intern – Mathematical Modeling and Industrial Solutions Group

Bengaluru, India

May – Nov 2016

- Built the initial blockchain invoice-discounting proof of concept for Mahindra & Mahindra on Hyperledger Fabric; the POC led to IBM securing the full project, the first blockchain project in India

Education

Indian Institute of Science

M.Sc. (Engg.) in Computer Science, Advisor: Prof. Shalabh Bhatnagar

Bengaluru, India

2014 – 2016

- **Thesis:** Stochastic Newton methods with enhanced Hessian estimation

NRI Institute of Technology

B.Tech. in Information Technology, Advisor: Dr. Viswanath Pulabaigari

Guntur, India

2008 – 2012

- Ranked 2nd in the department; first publication (k-nearest neighbor classification, 2010)

Expertise

- **Leadership:** Product ideation and strategy, technical direction, team building and mentoring, executive and customer engagement, research partnerships
- **AI:** Multi-agent systems, agentic AI, LLMs, reinforcement learning, knowledge graphs and ontologies, stochastic optimization, game theory
- **Platforms:** Kubernetes, GPU orchestration, multi-cloud infrastructure, Hyperledger Fabric
- **Engineering:** Python, Go, PyTorch, Neo4j, MongoDB

Awards

- | | |
|--|------|
| • IBM Research Spotlight Award | 2020 |
| • IBM Research India Code Reuse Championship | 2020 |
| • IBM Research India Travel Grant Award | 2020 |
| • Mentored project featured in Neuromatch Academy highlights | 2020 |
| • IBM Client Value Outstanding Technical Achievement Award | 2019 |
| • IBM Research Accomplishment Award | 2018 |
| • Top 1% of 155,190 candidates in GATE | 2014 |

Patents

- | | |
|---|-----------------|
| • Controlling Placement of Workloads of an Application within an Application Environment | US11805073B2 |
| • Data Quality Control | US11169976B2 |
| • Distributed Service Delivery System with Adaptive Autoscaling of Compute Resources | Filed |
| • Service Network Employing Reinforcement Learning-Based Global Load Balancing | Filed |
| • Multi-Objective Driven Refactoring of Monolith Application Using Reinforcement Learning | US20230177337A1 |
| • Decentralized Online Multi-Agent Visual Question Answering | US20220188362A1 |
| • Factchecking Artificial Intelligence Models Using Blockchain | US20220237477A1 |
| • Option-Based Distributed Reservation System | US20210158229A1 |

See full list of defensive publications on saikotireddy.com/research/

Publications

- [1] G. Shakya, **S. Danda**, S. Nath, P. Dayama, and S. S. K. Sajja, “Truthful and equitable lateral transshipment in multi-retailer systems”, in *26th European Conference on Artificial Intelligence, ECAI 2023*, IOS Press BV, 2023, pp. 2113–2120.
- [2] K. Sampath, S. Nishad, **S. Danda**, P. Dayama, and S. Sankagiri, “Inventory pooling using deep reinforcement learning”, in *2022 IEEE International Conference on Services Computing (SCC)*, IEEE, 2022, pp. 259–267.
- [3] P. Parnika, R. B. Diddigi, **S. Danda**, and S. Bhatnagar, “Attention actor-critic algorithm for multi-agent constrained co-operative reinforcement learning”, in *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2021, pp. 1616–1618, (**Joint first author**).
- [4] K. Narayanam, K. Sampath, K. Kumar, **S. Danda**, and P. Dayama, “Atomic execution of optimization transactions on permissioned blockchains”, in *2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, IEEE, 2021, pp. 1–9.
- [5] K. Sampath, **S. Danda**, K. Kumar, K. Narayanam, P. Dayama, and S. Sankagiri, “Spot collaborative shipping sans orchestrator using blockchain”, in *2020 IEEE International Conference on Blockchain (Blockchain)*, IEEE, 2020, pp. 371–378.
- [6] K. Sampath, P. Dayama, **S. Danda**, and V. D. Pandit, “Invoice reconciliation using blockchain”, in *INFORMS Annual Meeting*, Presentation, 2020.
- [7] K. Sampath, **S. Danda**, P. Dayama, and V. D. Pandit, “Collaborative shipping sans orchestrator using blockchain”, in *INFORMS Annual Meeting*, Presentation, 2020.
- [8] S. Nayak, C. A. Ekbote, A. P. S. Chauhan, R. B. Diddigi, P. Ray, A. Sikdar, **S. Danda**, and S. Bhatnagar, “Stochastic game frameworks for efficient energy management in microgrid networks”, in *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, 2020, pp. 116–120.
- [9] **S. Danda**, A. Saha, S. G. Tamilselvam, P. Agrawal, and P. Dayama, “Risk averse reinforcement learning for mixed multi-agent environments”, in *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2019, pp. 2171–2173.
- [10] R. B. Diddigi, **S. Danda**, K. J. Prabuchandran, and S. Bhatnagar, “Actor-critic algorithms for constrained multi-agent reinforcement learning”, in *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2019, pp. 1931–1933, (**Joint first author**).
- [11] R. B. Diddigi, **S. Danda**, K. Narayanam, and S. Bhatnagar, “A unified decision making framework for supply and demand management in microgrid networks”, in *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, IEEE, 2018, pp. 1–7, (**Joint first author**).
- [12] K. Chandramouli, K. J. Prabuchandran, **S. Danda**, and S. Bhatnagar, “Generalized deterministic perturbations for stochastic gradient search”, in *2018 IEEE Conference on Decision and Control (CDC)*, IEEE, 2018, pp. 5734–5739.
- [13] **S. Danda**, L. Prashanth, and S. Bhatnagar, “Improved Hessian estimation for adaptive random directions stochastic approximation”, in *2016 IEEE 55th Conference on Decision and Control (CDC)*, IEEE, 2016, pp. 3682–3687.
- [14] **S. Danda**, T. H. Sarma, P. Viswanath, and S. Raghava, “An improvement to k-nearest neighbor classifier”, in *3rd International Conference on Data Management (ICDM)*, 2010, pp. 314–324.

Teaching and Academic Service

- **Teaching Assistant**, Computational Neuroscience, Neuromatch Academy 2020
- **Teaching Assistant**, Algorithms and Programming, Indian Institute of Science 2016
- **Conference Reviewer**: IEEE ICDM (2019, 2020), ACC (2020), ACM CIKM (2019), IEEE CDC (2017)