SANJAY THAKUR

AI Systems Architect | Founder | Engineering Leader

+91 9900452668
sttsanjay@gmail.com

With over 700 citations in top ML conferences, I am a passionate technology leader with a proven track record in scaling teams, architecting high-performance systems, and driving business impact through AI and large-scale data platforms. Combines deep hands-on expertise in machine learning, real-time analytics, and distributed systems with strategic leadership, building and mentoring high-performing teams that deliver measurable results.

WORK EXPERIENCE

FOUNDER AND TECHNICAL HEAD

Widushi | June 2025 - Dec 2025

- **Directed R&D and full-cycle development** of the world's most evolved **handwritten answer evaluation tools (~95% human acceptance)** with a 4-person team, combining LLMS and image processing techniques with event-driven design.
- **Architected and deployed a scalable multi-platform solution** (mobile+web, backend, infrastructure & orchestration) via React, React Native (Expo), PostgreSQL, FastAPI, Terraform, hosted on GCP+AWS, with a focus on **enterprise grade reliability.** It achieved consistent high-quality output across 100K+ evaluated answers.
- **Scaled data pipelines & user systems** to support more than **1,000 students** and secured **3 B2B organizational clients,** establishing data definitions, key metrics, and lean governance frameworks for consistent analytics.
- **Technical Skills:** LLMs, Image Processing, Full-Stack, React JS, React Native (Expo), PostgreSQL, FastAPI, GCP, AWS.
- **Managerial Skills:** Team Leadership, Product Strategy, Full-Cycle R&D, Lean Production/Shipping.

VP, ENGINEERING

Curium | May 2023 - June 2025

- **Led an engineering team of 10+** across ML Vision, data infrastructure, and platform, managing complex dependencies and defining measurable outcomes for 1000+ deployments across 100+ of locations.
- **Championed GDPR/CCPA-aligned** data privacy practices and audit compliance across analytics pipelines and distributed edge deployments.
- **Architected data pipelines and scaled distributed workflows** for real-time analytics and reliability reporting across systems reducing operational overhead from 20+ people to 2. It further improved system uptime from 95% to 99.99% and reduced deployment latency by 60%.
- **Recruited and developed engineering talent**, scaling the team from 4 to 11, and implemented performance management frameworks to foster ownership and autonomy. Further, mentored them in agile methodologies and architecture design.
- **Established MLOps practices** using GitLab CI/CD, Docker, and Terraform,
- **Technical skills:** ML (PointPillars), Cloud, ELK, Observability & Monitoring, Backend (FastAPI), Database (PostgreSQL, MongoDB), Terraform
- **Managerial skills:** Agile leadership, cross-functional collaboration, strategic planning, team mentoring

CO-FOUNDER,CTO

Markovian AI (acquired)| 2020 - 2022

- **Led a 3-member core team** as the primary developer to build a data semantic search platform using BERT models, scaling to serve 5 B2B enterprises across 3 countries while managing distributed teams.
- **Optimized query latency** from hours to <30 seconds by implementing Redis caching and custom indexing for petabyte-scale analytics using Apache Cassandra.
- **Architected high-throughput ETL pipeline** (>100MB/sec) using Kafka, Apache Spark & Airflow for real-time pattern detection, while mentoring 4 engineers in distributed systems while, reducing infrastructure costs by 40% while maintaining 99.99% uptime.
- **Technical skills:** Data ETL (Kafka, Spark, Airflow), Backend (FastAPI), Search Systems (Elasticsearch) .
- **Managerial skills:** Product strategy, stakeholder management, technical roadmapping, startup leadership

APPLIED ML RESEARCHER

Deeplite Inc | Sept 2019 - Mar 2020

- **Developed efficient ML models** achieving 5x size reduction using quantization and matrix decomposition, while leading knowledge sharing sessions for team.
- **Implemented ONNX**-based knowledge distillation based pipeline compatible with TensorFlow and PyTorch, coordinating with embedded systems team.
- **Built a custom CI/CD framework** that supports ARM64 based builds for supporting efficient neural networks to be deployed on CCTV cameras. Eliminate manual workflows that used to span 3-4 days.
- **Technical skills:** Python, TensorFlow, PyTorch, ONNX, ARM64.

PERSONAL SKILLS

Creativity
Team building
Communication
Problem Solving
Leadership

EDUCATION

MS CS, AI SPECIALISATION

MILA & Mobile Robotics Lab | 2017 - 2019 Publication *(ICRA 2019)*

MS CS, THESIS (LINK)

McGill University, Montreal | 2016 - 2019

BACHELOR OF TECHNOLOGY, COMPUTER SCIENCE

Malaviya National Institute of Technology, Jaipur | 2012 - 2016