Subhadip Mitra

Engineering Leader & AI Systems Architect

Singapore   contact@subhadipmitra.com   LinkedIn   GitHub

Senior Engineering Leader with 15+ years bridging AI research and enterprise delivery. Leading Google Cloud's Data & Analytics practice for Southeast Asia while driving innovations in LLM inference and multi-agent systems.

Experience

Head of Data & Analytics, Southeast Asia

Google Cloud - Professional Services Organization   2021–Present

- Built D&A practice from 0 to 1 across 6 countries; 97% CSAT
- Site Lead overseeing delivery governance across 7 PSO practices (Data Analytics, AI/ML, Infrastructure, Security, Enterprise Architecture, AppDev, Delivery Management)
- Pioneered agentic AI adoption across 7 practices and 6 JAPAC sub-regions with SDKs, agent catalog, and governance frameworks
- Spearheaded rescue operations for at-risk enterprise accounts, recovering strategic customers
- Partner with Sales leadership on strategic pursuits; collaborate with Product Engineering on platform roadmap
- Published 5 technical disclosures (UPIR, FTCS, ARTEMIS, ETLC, CatchMe) on AI/distributed systems

Principal Engineer - Data & Analytics Transformation

Standard Chartered Bank   2019–2021

- Delivered Self-Service ML Platform reducing model development from 6 months to 1 week
- Modernized MarTech infrastructure driving 30% increase in customer acquisition

Principal Data Engineer / Solution Architect

Think Big Analytics (Teradata)   2017–2019

- Engineered 5 data lakes processing 1.2 PB/hour with 20% optimization
- Built real-time fraud detection reducing false positives by 60%

Software Engineering & Technical Leadership

Microsoft, UTU & Others   2010–2017

Education

MBA, Business Analytics

BITS Pilani   2021–2023

MTech, Software Systems

BITS Pilani   2017–2020

Research & Innovation

LLM Inference Efficiency (2025)

Speculative decoding, custom Triton kernels, KV-cache compression. 8.1x speedup, 88% peak bandwidth on A100.

AI Metacognition Toolkit (2025)

Activation-level detection of sandbagging and deception in LLMs. 90-96% accuracy with steering vectors.

UPIR - Distributed Systems Synthesis (2025)

Neuro-symbolic framework for automated distributed system generation. 274x speedup, 60% latency reduction.

CatchMe - Intelligent Trust Engine (2025)

Agentic AI for enterprise trust decisions. APLS pattern learning + cascade routing. 86% cost reduction, sub-50ms latency.

Spark LLM Eval - Distributed Evaluation Framework (2025)

Enterprise-scale LLM evaluation on Apache Spark. Statistical rigor with bootstrap intervals, multi-provider support (OpenAI, Anthropic, Gemini), Delta Lake integration.

Steering Vectors for Agent Behavior (2025)

Runtime control of LLM agent behaviors through activation steering. 65% uncertainty detection while maintaining 100% confidence on factual queries.

Publications

- UPIR: Automated Synthesis and Verification of Distributed Systems (2025)
- ETLC: Context-First Data Processing for GenAI (2025)
- Field-Theoretic Context System (FTCS) (2025)
- ARTEMIS - Adaptive Multi-agent Debate Framework (2025)
- OConsent: Blockchain Privacy & Consent Protocol (2021)

Skills

Leadership: Enterprise Architecture, Digital Transformation, AI & Data Strategy, C-Suite Advisory

Data & AI: Petabyte-Scale Platforms, Apache Spark/Kafka/Flink, Delta Lake, LLMs & Multi-Agent Systems, RAG & Vector Databases, PyTorch, Vertex AI

Cloud & Engineering: Google Cloud Platform, BigQuery, Python/SQL/Scala, Kubernetes, Distributed Systems