

Subhadip Mitra

Engineering Leader & AI Systems Architect

Singapore contact@subhadipmitra.com +65 82501776 [LinkedIn](#) [GitHub](#)

Senior Engineering Leader with 15+ years building teams, frameworks, and systems that turn Data and AI from research to production. Head of Data & Analytics for Google Cloud in Southeast Asia - a practice built from zero, delivering enterprise transformation across 7 countries. "Player-Coach" leading petabyte-scale platforms while driving published research (5 technical disclosures, 8 packages on PyPI/Maven, open-source AI safety tools).

Experience

Head of Data & Analytics, Southeast Asia — Site Lead, PSO Southeast Asia

Google Cloud - Professional Services Organization 2021–Present

- Built D&A practice from 0 to 1 across 6 countries; Site Lead overseeing 7 PSO practices' delivery in SEA; 97% CSAT, 100% revenue target attainment
- Delivered first-of-kind solutions including GenAI-powered reconciliation (replicated across JAPAC), ML platform migrations (30K+ notebooks), petabyte-scale modernizations
- Pioneered agentic AI adoption across 7 practices and 6 JAPAC sub-regions with SDKs, agent catalog, architecture discovery (100M+ node graphs), and governance frameworks
- Led cross-practice rescue operations for at-risk enterprise accounts, recovering strategic customers and converting platform exits into long-term partnerships
- 5 Google Technical Disclosures (UPIR, FTCS, ARTEMIS, rule generation, cost-benefit routing) plus ETLC whitepaper; 8 packages on PyPI/Maven Central

Principal Engineer - Data & Analytics Transformation

Standard Chartered Bank 2019–2021

- Built data & analytics platform serving 11 markets, 100+ systems, and 1200+ users
- Delivered Self-Service ML Workbench reducing model deployment from months to weeks
- Built credit risk models over 15K+ entities using news/social signals, reducing losses by \$5M

Principal Data Engineer / Solution Architect

Think Big Analytics (Teradata) 2017–2019

- Designed 5 data lakes processing 1.2 PB/hour and 40K daily files
- Engineered real-time platform processing 2.5M events/second, improving Ad campaign responsiveness by 80%
- Built ML fraud detection: 60% fewer false positives, \$3M savings

Software Engineering & Technical Leadership

Microsoft, Truckaibus (Founder), UTU 2010–2017

- Microsoft (2010-2014): Windows Kernel development (Windows 7/8, Server 2012 R2), Azure ML, CDN optimization
- Truckaibus (2014-2016): Founded B2B commercial vehicle marketplace - 15 cities, 25+ OEM/bank partnerships
- UTU Singapore (2016-2017): Led Thailand development; bank integration; payment/rewards systems

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

- [Automated Rule Generation for Tiered Systems \(2026\)](#)
- [Predictive Cost-Benefit Routing for Risk Decisioning \(2026\)](#)
- [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