Subhadip Mitra

Engineering Leader & AI Systems Architect

Singapore | +65 82501776 | contact@subhadipmitra.com
LinkedIn | GitHub | Website

Last updated: February 21, 2026

Professional Summary

Senior Engineering Leader with 15+ years building the teams, frameworks, and systems that turn Data and AI from research to production. Currently Head of Data & Analytics for Google Cloud in Southeast Asia - a practice built from zero, delivering enterprise Data and AI transformation across 7 countries. Dual track as "Player-Coach": leading petabyte-scale data platforms and multi-agent systems for Fortune 500 clients, while driving innovation through published research (5 technical disclosures, 8 published packages on PyPI and Maven Central, plus open-source AI safety tools including sandbagging detection and activation steering). Member of Google Cloud delta, architecting solutions at the intersection of applied AI and enterprise scale.

Professional Experience

Google Cloud - Professional Services Organization: Head of Data & Analytics, Southeast Asia — Site Lead, PSO Southeast Asia *January 2021 – Present*

Dual-track role combining technical innovation leadership with regional delivery management. Built Google Cloud's Data Analytics practice for Southeast Asia with delivery scope across JAPAC, while serving as Site Lead overseeing cross-practice operations in SEA. Member of delta - Google Cloud's innovation and transformation team architecting enterprise AI solutions at scale.

Strategic Leadership & Delivery:

- Built Data Analytics practice for Southeast Asia from 0 to 1 across 6 countries. Serve as Site Lead overseeing all 7 PSO practices' delivery in SEA, owning utilization and CSAT metrics (97%) and contributing to 100% annual revenue target attainment.
- Delivered first-of-kind solutions including GenAI-powered reconciliation framework for a major airline (now replicated across JAPAC), large-scale ML platform migrations (30K+ notebooks), and petabyte-scale data platform modernizations for Fortune 500 clients across financial services, telcos, and consumer electronics.
- Led cross-practice rescue operations for at-risk enterprise accounts with multi-million dollar project values, recovering strategic customers and converting potential platform exits into long-term partnerships.
- Pioneered agentic AI adoption across all 7 PSO practices and 6 JAPAC sub-regions. Built SDKs, agent catalog, architecture discovery tools (100M+ node graph modeling), automated pipeline generation, and governance frameworks that reduced delivery costs.
- Built Data Strategy competency from 0, delivering 8-figure pursuit value across 14 strategic pitches in Asia Pacific. Partner with C-level stakeholders (CTOs, CDOs) to define data modernization and AI transformation roadmaps.

Technical Innovation & Research:

- 5 Google Technical Disclosures (UPIR, FTCS, ARTEMIS, rule generation for tiered systems, cost-benefit routing for risk systems) plus the ETLC whitepaper on context-first data processing for GenAI. 8 open-source packages published on PyPI and Maven Central.
- Industry-agnostic agentic AI for enterprise trust decisions. APLS self-learning + cascade routing achieving 86% cost reduction, sub-50ms latency.

Standard Chartered Bank: Principal Engineer - Data & Analytics Transformation *January 2019 – January 2021*

Led design and development of retail bank's data & analytics platform serving 11 markets, 100+ systems, and 1200+ users.

- Developed self-service ML Workbench reducing model deployment time from months to weeks
- Architected MarTech strategy driving 30% increase in customer acquisition through data-driven personalization
- Created credit risk models over 15,000+ named entities leveraging news trends and social signals, reducing potential losses by $5M
- Defined enterprise data strategy including third-party data governance, privacy frameworks, and cloud adoption roadmap

Think Big Analytics (a Teradata company): Principal Data Engineer / Solution Architect *January 2017 – January 2019*

Architected enterprise-scale data solutions for Fortune 500 clients across APAC.

- Designed 5 global data lakes with ETL pipelines handling 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 system achieving 60% fewer false positives and 25% higher detection rates, resulting in $3M savings
- Built and managed large-scale Hadoop clusters (300+ nodes) for banks and telcos across JAPAC

Microsoft, Truckaurbus (Founder), UTU: Software Engineering & Technical Leadership *January 2010 – January 2017*

Progressive advancement through software engineering, entrepreneurship, and technical leadership across systems development, marketplace platforms, and payments infrastructure.

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

Research & Open Source Engineering

Spark LLM Eval - Distributed Evaluation Framework
Distributed LLM evaluation framework built on Apache Spark for enterprise-scale model assessment. Addresses the gap in evaluating LLMs at scale with statistical rigor, integrating seamlessly with Databricks infrastructure.
GitHub · Blog

LLM Inference Efficiency Research
Research implementations addressing the fundamental bottleneck in LLM inference: memory-bandwidth constraints rather than compute limits. Explores acceleration through speculative decoding, custom GPU kernels, and quantization strategies.
GitHub · GitHub
Google Technical Disclosure - Pending

AI Metacognition Toolkit
Activation-level detection of sandbagging, deception, and situational awareness in LLMs. Linear probes achieve 90-96% accuracy across Mistral, Gemma, and Qwen models. Published on PyPI.
PyPI · GitHub · Docs · Blog

Steering Vectors for Agent Behavior Control
Runtime control of LLM agent behaviors through activation steering vectors - modifying model outputs at inference time without retraining. Demonstrates more calibrated control than traditional prompting approaches with LangChain integration.
GitHub · Blog

Publications & Technical Disclosures

Automated Rule Generation for Tiered Systems Using Multi-Stage Failure Learning
Google, Technical Disclosure Commons, January 2026

Predictive Cost-Benefit Routing for Multi-Tier Risk Decisioning Systems
Google, Technical Disclosure Commons, January 2026

UPIR: Automated Synthesis and Verification of Distributed Systems
Google, Technical Disclosure Commons, November 2025

ETLC: A Context-First Approach to Data Processing in the Generative AI Era
Google Cloud, May 2025

Field-Theoretic Context System (FTCS)
Google, Technical Disclosure Commons, May 2025

ARTEMIS - Adaptive Multi-agent Debate Framework
Google, Technical Disclosure Commons, January 2025

Data Monetization Strategy for Enterprises
BITS Pilani, December 2023

OConsent: Open Consent Protocol for Privacy and Consent Management with Blockchain
BITS Pilani, December 2021

Education

MBA, Business Analytics — Birla Institute of Technology and Science, Pilani
MTech, Software Systems — Birla Institute of Technology and Science, Pilani

Technical Skills

Technology Leadership & Strategy: Enterprise Architecture, Digital Transformation, AI & Data Strategy, C-Suite Advisory, Innovation Leadership, Strategic Planning

Data Engineering & Architecture: Data Pipelines, Real-Time Processing, Data Mesh & Fabric, Data Governance, Apache Spark, Delta Lake, Apache Kafka, Apache Iceberg

Generative AI & Machine Learning: Multi-Agent Systems, Large Language Models, RAG Architecture, Vector Databases, PyTorch, LangChain, LangGraph, LlamaIndex, Google ADK, MCP, A2A Protocol, MLflow, LLMOps

Cloud Platforms & Infrastructure: Google Cloud Platform, BigQuery, Vertex AI, Dataproc, Cloud Composer, GKE, Terraform, Kubernetes

Programming & Development: Python, SQL, Scala, Triton, CUDA, Algorithm Design, Formal Verification, Program Synthesis, Distributed Systems