Subhadip Mitra

Data & AI Executive

Singapore · contact@subhadipmitra.com · LinkedIn · GitHub

I build the practices, teams, and systems that turn AI from research into enterprise reality. 15+ years driving enterprise transformation across APAC as a Data & AI executive. Head of Data & Analytics, Google Cloud Southeast Asia – built the practice from zero to ~$X0M in delivery value, overseeing ~$X0M regional portfolio as Site Lead; 8 technical disclosures and peer-reviewed research published at ICLR.

Experience

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

- Built D&A practice from zero to ~$X0M delivery value across 6 countries; as Site Lead, oversaw ~$X0M regional portfolio across all 7 SEA practices – 97% CSAT and 100% revenue targets
- Created Practice's Data Strategy competency ($X0M+ pursuit pipeline in 2 years). Rescued strategic account facing platform exodus, converting to $X0M+ 3-year commitment
- Delivered first-of-its-kind ML notebook migration (30K+) via custom multi-agent system, GenAI reconciliation for a major airline (replicated across JAPAC), and petabyte-scale modernizations for Fortune 500 clients
- Designed and launched JAPAC's Agent Program Office (APO), operationalizing AI agents across 5 practices and 6 sub-regions. Made agentic assessment mandatory for all Deal Reviews, launched 7 pilot projects. Targeting 10% efficiency across 120+ person org

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 with 500+ production models, reducing deployment from 6 months to 1 week
- Built credit risk models over 15K+ entities using news/social signals, reducing losses by $XM

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, $XM savings

Entrepreneurship & Technical Leadership, *Truckaurbus (Founder & CTO) / UTU Singapore* 2014–2017

Senior SDE, *Microsoft* 2010–2014

Education

MBA, Business Analytics
BITS Pilani 2021–2023

MTech, Software Systems
BITS Pilani 2017–2020

Research & Innovation

LLM Inference Efficiency (2025)
Speculative decoding (2-3x speedup) and custom Triton kernels (8.1x faster, 88% peak A100 bandwidth). Google Technical Disclosure pending.

AI Metacognition Toolkit (2025)
Sandbagging and deception detection in LLMs via activation probes. 90-96% accuracy; steering vectors reduce sandbagging 20%. Published on PyPI.

UPIR - Distributed Systems Synthesis (2025)
Neuro-symbolic framework for automated distributed system generation. 274x speedup, 60% latency reduction. Google Technical Disclosure.

Steering Vectors for Agent Behavior (2025)
Runtime control of LLM agent behaviors via activation steering. 65% uncertainty detection, 100% confidence on factual queries.

Publications

- Quality-Diversity Evolution for LLM Safety Vulnerabilities (2026, *ICLR 2026 Workshop AIWILD*)
- Field-Theoretic Memory for AI Agents (2026, *arXiv*)
- 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)
- ARTEMIS - Adaptive Multi-agent Debate Framework (2025)

Skills

Strategy & Leadership: Practice Building & P&L, Executive Advisory, Data Strategy & Monetization, Go-to-Market Partnership, Talent Development, Org Design

Technical Domains: Enterprise AI/ML Platforms, Agentic AI Systems, Data Platform Modernization, Google Cloud Platform, Distributed Systems

Notable Projects

CatchMe - Intelligent Trust Engine (2025)
Industry-agnostic agentic AI for enterprise trust decisions. Self-learning cascade routing with 86% cost reduction and sub-50ms latency.

LLMConsent (2024 - Present)
Privacy-preserving consent protocol for LLM training data. Decentralized registry with cryptographic verification and real-time opt-out enforcement.

SMPP Core (2024 - 2025)
Java 21 SMPP protocol implementation with virtual threads. 1.8M PDU decodes/sec, 1.5M encodes/sec. Published on Maven Central.