Srihari Medarametla

medarsri@gmail.com | +91-9449051271 | linkedin.com

Principal Architect | AI Systems | Governance | Regulated Environments | Enterprise Architecture

PROFESSIONAL SUMMARY

- Works across **AI engineering, AI risk governance, data governance** and **C-Suite Advisory**. Most of that work has been **production AI systems** that have to hold up under real **regulatory frameworks**.
- Led architecture and deployment of **multi-agent AI pipelines** replacing largely manual compliance workflows. Built end-to-end **AI governance frameworks** aligned to **NIST AI RMF** and **DAMA-DMBOK** across portfolio of production AI models, resolving regulatory findings by implementing **bias controls, explainability** and **model risk management**.
- Led client-facing pre-sales engagements, **winning competitive RFPs** through end-to-end POC-to-commercial proposal delivery.

CORE COMPETENCIES

- **AI Architecture and Engineering:** Multi-agent systems (LangGraph, ReAct, HITL), Agentic AI, GPT-4o, GPT-4.1, Azure OpenAI, Advanced RAG architectures (HyDE, Hybrid Search, Cohere Rerank, Parent-Child Chunking), XGBoost, SHAP, Fairlearn, MLflow, LangSmith, FastAPI, ML Algorithms, NLP, Python
- **AI and Data Governance:** Governance/Regulatory Guardrails (Schema validation, Determinism, IAM, Data Privacy, Data Residency, Audit Trail, Security, Model Risk Management), NIST AI RMF, DAMA-DMBOK v3, EU AI Act, FCA, DORA, GDPR, SR 11-7, DPDP 2023, Collibra, Immuta ABAC, Azure Purview, Federated Learning, Responsible AI, Explainable AI
- **Cloud, Data and Engineering:** Azure OpenAI, Azure AKS, Azure Event Grid, Azure APIM, Azure DevOps, Docker, Databricks Unity Catalog, Delta Lake, PGVector, Azure AI Search, Data Mesh, MDM, LangChain LCEL, SQL, Power BI
- **Client Advisory and Pre-Sales:** RFX management, Solution architecture, POC design and validation, Value-based pricing, C-suite advisory workshops, Technical proposal authoring, Executive oral presentations, Product Management (0-to-1), Design Thinking, User Research, UX Design, Progressive Disclosure, PRD Authoring, Product-Market Fit, ROI Modelling, Figma, Miro
- **Enterprise Programs and Leadership:** Multi-phase-year-geo Program Delivery, Requirements Synthesis at Enterprise Scale, Senior Stakeholder Orchestration (CDO, CRO, CCO, PPEDs), Executive and Board Reporting, Cross-functional Team Leadership, Interdepartmental Alignment, SAFe / Agile / Scrum

PROFESSIONAL EXPERIENCE

End-to-End Product Architect | (Independent Venture, early stage) May 2026 – Present

An early-stage solo venture, still testing the concept and the market. I run an AI-orchestrated content studio with three independent lines off one shared AI pipeline (LLMs for scripting, AI video generation, plus voice and image tooling): a micro-drama streaming platform (One Minute Dramas), a comedy channel (Pappadamix), and a spec-advertising studio (AdChaatwala). I own creative origination, architecture, build and publishing across all the **three different product lines**.

- **AI orchestration:** One shared pipeline for scripting, image referencing, video generation, voice synthesis and edit, directed across all three genres and audiences.
- **IP creation:** Authored creative base like episode outline for micro-drama vertical, original animated format and script for the second, both copyrightable, self-originated IP. The third line runs on an unsolicited speculative ad model, governed by disclaimer and trademark-avoidance discipline.
- **Architecture and hands-on build:** Developed a Next.js PWA with HLS streaming (Bunny.net), with a planned Razorpay payments core as the commercial backbone for the drama vertical; the two newer lines run YouTube-native for now.

- **Compliance and risk posture, by line:** Planned for a prepaid-instrument, GST/OIDAR, DPDP 2023 obligations plus deferred-vs-recognised revenue into the drama platform's data model itself; a standing disclaimer and trademark-avoidance framework for the spec-advertising line, carried through descriptions and on-screen text.
- **Unit economics, Go-to-market:** Performed per-line unit-economics models isolating the true scaling cost before committing further spend. Created brand identity, channel assets and launch collateral.

Stack: Next.js, Supabase (PostgreSQL, RLS), Bunny.net Stream, PWA, Claude/Gemini, Kling AI 3.0, ElevenLabs, Nano Banana, VN Editor

Project Director | C-DAC

Jul 2025 – Jan 2026

Led program governance, stakeholder coordination, and AI strategy delivery for India's **National Supercomputing Mission** — a government-commissioned, multi-crore, multi-year, exascale HPC-AI national infrastructure program spanning multiple central R&D and academic institutes and government ministries.

- **National Infrastructure Strategy:** Orchestrated multiple domain executive stakeholders (Director level and above) across central R&D institutes and ministries to author key strategic pillars of the **Detailed Project Report (DPR)** for India's exascale HPC infrastructure with roadmap.
- **Sovereign Data Governance Framework:** Drafted a comprehensive data governance framework for the NSM national multi-tenant **HPC-AI Data Center** hosting data workloads for Defence, ICAR, ICMR, IMD, and IITs. Defined 4-tier data classification (Open to Classified Sovereign) governing storage architecture, network zoning, and secure compute isolation. DPDP Act 2023, CERT-In, and MeitY compliant.
- **AI Onboarding Standards:** Standardised **AI protocols** integrating FAIR data standards, training data provenance tracking, inference controls, model registry, and privacy-preserving Federated Learning governance for sensitive data.
- **Program Planning and Governance:** Defined program milestones, workstream dependencies, and delivery roadmap across concurrent HPC infrastructure. Maintained live risk and issue register, facilitated fortnightly program review cadences with Director-level stakeholders using structured status reports and RAG dashboards. **Coordinated a distributed stakeholder landscape** spanning DST, MeitY, ICAR, ICMR, IMD, IUCAA, and IITs.
- **AI/ML Health Informatics:** Research project on **Digital Twin for Disease Modelling** and Drug Discovery integrating ML and knowledge graphs on multi-omics data.

Stack: AI research: NotebookLM, DPDP Act 2023, CERT-In, FAIR principles, Federated Learning, MeitY and DST policy frameworks

AI & Data Architect II | EY (Ernst & Young)

Jun 2019 – Jun 2025

KYC and AML Compliance Automation - Lead AI Architect

Replaced a 9-day largely manual KYC/AML process with an **autonomous five-agent pipeline**. Designed for FCA SM&CR compliance with HITL interrupts.

- **Architecture decision:** ReAct agent pattern selected over a deterministic workflow because data-gathering steps adapt per applicant type (document sets, corporate vs individual), which needs LLM-directed tool sequencing. Parallel fan-out (Agents 2 and 3) eliminated sequential dependency. HITL, architecturally non-bypassable.
- **Governance design:** All five agents emit structured SHAP per-decision explainability outputs to an append-only Cosmos DB WORM audit trail. GPT-4o case synthesis constrained to evidence from pipeline state only, with no external knowledge injection.
- **Outcomes:** Processing time from 9 days to under 25 minutes for standard-risk cases (72% of volume). Sanctions false positives reduced 38% vs. legacy rule-based system. Compliance officer manual caseload reduced 64%.

Stack: LangGraph, GPT-4o, Azure Document Intelligence, Refinitiv World-Check, OFAC SDN, Azure Cosmos DB WORM, Azure AKS, LangSmith, FastAPI, Docker, Python

Regulatory Change Impact Management - Lead AI Architect

Designed an **event-triggered workflow monitoring multiple regulatory bodies** (FCA, EBA, DORA, BCBS, PRA, ESMA, ECB, RBI, MAS), automatically progressing from publication to drafted policy amendments within 30 mins of publication.

- **Architecture decision:** Chose a deterministic DAG over ReAct here because all five processing steps/nodes (Ingest, Triage, Impact, Drafting) are invariant across every document type, so the workflow is correct by design and produces the same, pre-defined DORA Article 11 audit record for every document.
- **Outcomes:** Regulatory response time reduced from 4–6 weeks to 3–4 working days. Policy impact coverage from 60% to 91%. 1,200 analyst-hours saved per quarter. DORA Article 11 gap resolved.

Stack: LangGraph, GPT-4.1, Azure AI Language, PGVector, Collibra Policy Center API, Azure Event Grid, Azure AKS, PostgreSQL, Python

Advanced Q & A Knowledge Assistant - Lead AI Architect

Designed a **multi-stage advanced RAG pipeline** for a 4000+ document knowledge base serving four RBAC-controlled user tiers (Customer, Advisor, Agent, Compliance). RBAC enforced at the lowest-level retrieval function, not post-retrieval, so FCA-restricted documents are never surfaced to lower access tiers regardless of query construction.

- **Advanced RAG design:** HyDE expansion resolves vocabulary gap between user queries and regulatory document language. Multi-query generation (4 variants). Cohere Rerank cross-encoder (top-20 to top-5) outperforms cosine similarity for legal text. Parent-child chunking (512-token child for retrieval precision; 2048-token parent sent to LLM for complete context). Groundedness validated by a separate independent LLM judge call.
- **Outcomes:** Query response time from 4–8 minutes to under 6 seconds. First contact resolution from 62% to 81%. RAGAS Faithfulness score 0.91. FCA Consumer Duty citation compliance: 100%.

Stack: GPT-4o, Azure AI Search, Cohere Rerank v3, LangChain LCEL, Azure Cosmos DB, RAGAS, Azure AI Content Safety, Presidio, Python

AI Governance and Data Trust Program - AI & Data Architect

Designed and delivered an **enterprise AI governance framework** across 40+ production AI models, responding to active FCA supervisory findings on algorithmic bias in credit scoring.

- **Framework design:** Aligned to NIST AI RMF with GOVERN function implemented first, which set up risk classification, accountability mapping, and documentation standards. Integrated governance gate into CI/CD pipeline, blocking all AI model deployments until governance conditions are met.
- **Controls implemented:** Fair learn MetricFrame bias detection (DPR >= 0.80, equalised odds <= 0.05 as deployment gates). SHAP TreeExplainer per-model explainability reports auto-generated. Azure Purview automated discovery, cataloguing, sensitivity labelling, custom lineage tracking. Immuta ABAC automated data access, enabling self-serve and GDPR. Art 5(1)(b) purpose limitation at access time itself. Quarterly AI Risk Review via Power BI.
- **Outcomes:** FCA supervisory findings resolved. Fairlearn DPR audit identified disparity on protected characteristics (<0.8), were remediated prior to deployment. Missing SM&CR owner were assigned. 40+ models governed.

Stack: NIST AI RMF, MLflow, Azure Purview, Immuta, Fairlearn, SHAP, EU AI Act, FCA, SR 11-7, DAMA-DMBOK, Python

AI Governance pre-sales pitch - US Wealth Management - AI Solutions Architect (Pre-Sales)

Led the **client visit demo for a large US Wealth Management firm** requiring AI governance and regulatory compliance architecture.

- **Pre-sales approach:** Client discovery workshop through to a live POC that showed bias detection, explainability, model risk management, and Immuta ABAC.
- **Delivered:** Full NIST AI RMF architecture proposal with bottom-up WBS, 3-phase implementation roadmap, and value-based pricing anchored on quantified SEC, FINRA, Reg BI, regulatory risk reduction.
- **Outcomes:** Progressed to RFP phase and won against 2 competitors. Engagement contracted across first phase.

Other Notable Engagements

- **Trusted Data Platform Initiative:** Led the Experience and Data Management workstreams for EY's CDO commissioned enterprise data governance and consumption platform, pulling fragmented, siloed data capabilities into a single governed ecosystem. Helped define a layered **Conceptual Architecture** (Portal - Consumption - Platform Services - Intelligence - Supply Chain). Authored **PRDs** via stakeholder consensus across CDO, Engineering, DG, BI, Compliance, IT Architecture, and LOBs. Produced a **Function Map** that gave the different teams one shared vocabulary and settled a lot of the arguments about who owned what. It doubled as the UX research scope and the delivery roadmap. Delivered end-to-end **UX design**: user research, persona profiles, journey maps, and hi-fidelity Figma wireframes covering Data Discovery, Marketplace, self-service Request Workflow, and Data Quality. Presented **product vision to CDO** and program board. **Outcomes:** program advanced from Ideate to Build phase.
- **Data Governance Project:** Designed a unified architecture on 3-platform tier **Collibra, Immuta, Databricks** mapped to GDPR, FCA and Model Risk standards for a financial services firm. End-to-end lineage ingestion across data integration pipelines alongside data quality gates between Medallion zones. Immuta ABAC policy engine to enforce real-time PII masking, automated purpose limitation safeguards, and self-serve data access. **Outcomes:** Enterprise Data Governance Platform architectural design delivered.
- **Manufacturing Defect Detection:** Implemented a ML solution, **predicting defects on kitchen slabs** before the costly final quality inspection step by analysing **hundreds of sensor readings across manufacturing stages** (mixing, degassing, casting, curing) to produce a defect probability score per slab. The PLC/SCADA system generated ~100 GB of raw sensor logs daily across multiple lines and shifts. The primary **architectural challenge was data engineering** — transforming

large flat-file buffers into a clean, analysis-ready format. **Outcomes:** Handled using Apache Dask. XGBoost AUC 0.918. Defect detection precision: 0.847. False alarm rate: 8.3%.

- **Smart Meeting Scheduler Add-in:** Designed and implemented an intelligent meeting scheduler that **minimises unnecessary cross-location travel and carbon footprint** for a globally distributed organisation. Given a meeting type (strategy, workshop, 1:1 etc), the system determines the optimal format, e.g., recommending face-to-face in a boardroom where role and meeting type justify it, or a MS-Teams call where it does not. Automatically queries Outlook calendars and working hours via Microsoft Graph API, resolves attendee roles and office locations from Azure AD metadata, HRD data, and audits local facility inventories (Telepresence, Immersive Video, Interactive Whiteboards, Video Conferencing) for real-time availability. **Outcomes:** Outputs a ranked set of **recommendation cards** – No Travel, Least Travel with best experience, Most Optimised, and Max Travel for face-to-face etc, each with time slot, optimal venue, estimated travel cost, and carbon savings. Sits on outlook toolbar as an Add-in.
- **Pre-Sales Advisory:** Solution architecture, effort estimation, WBS, proposal authoring, and executive presentations for more than **20 engagements** across Banking, FinTech, Risk, and AI portfolios over the 6-year period.

Principal Consultant | Wipro

Jun 2018 – Jun 2019

Applied **AI/ML and NLP on enterprise documents** (topic modelling, vector comparison) in prediction and automation for Document Intelligence use cases. Led related pre-sales consulting work.

Research Scholar | Indian Institute of Science

Aug 2012 – Jun 2018

Applied advanced quantitative methods to large-scale institutional data — OLS regression, Negative Binomial regression, fixed/random-effects panel models, cross-lagged Structural Equation Modelling (SEM), and patent network graph analysis. **That grounding in applied statistics is what I still lean on for current AI modelling, governance and explainability work**.

Research focus: Technology Transfer and Innovation Ecosystems. Covered technology innovation evolution and transfer, patent analytics, IP governance across Indian University and R&D systems.

Technology Architect | Infosys

Oct 2006 – Jul 2012

Managed large software testing programs at client sites in India and the US. Team size 8 to 28. Technical SME, Pre-sales support and RFP management.

Senior Technical Associate | Tech Mahindra

Jul 2002 – Sep 2006

Full-stack product development for Telecom and Commercial Banking platforms.

EDUCATION

Ph.D. | Applied Statistics and Innovation Management | Indian Institute of Science (IISc), Bangalore

M.S. | Telecommunications and Software Engineering | Birla Institute of Technology and Science (BITS), Pilani

B.Tech. | Electronics and Communication Engineering | College of Engineering (CET), Trivandrum