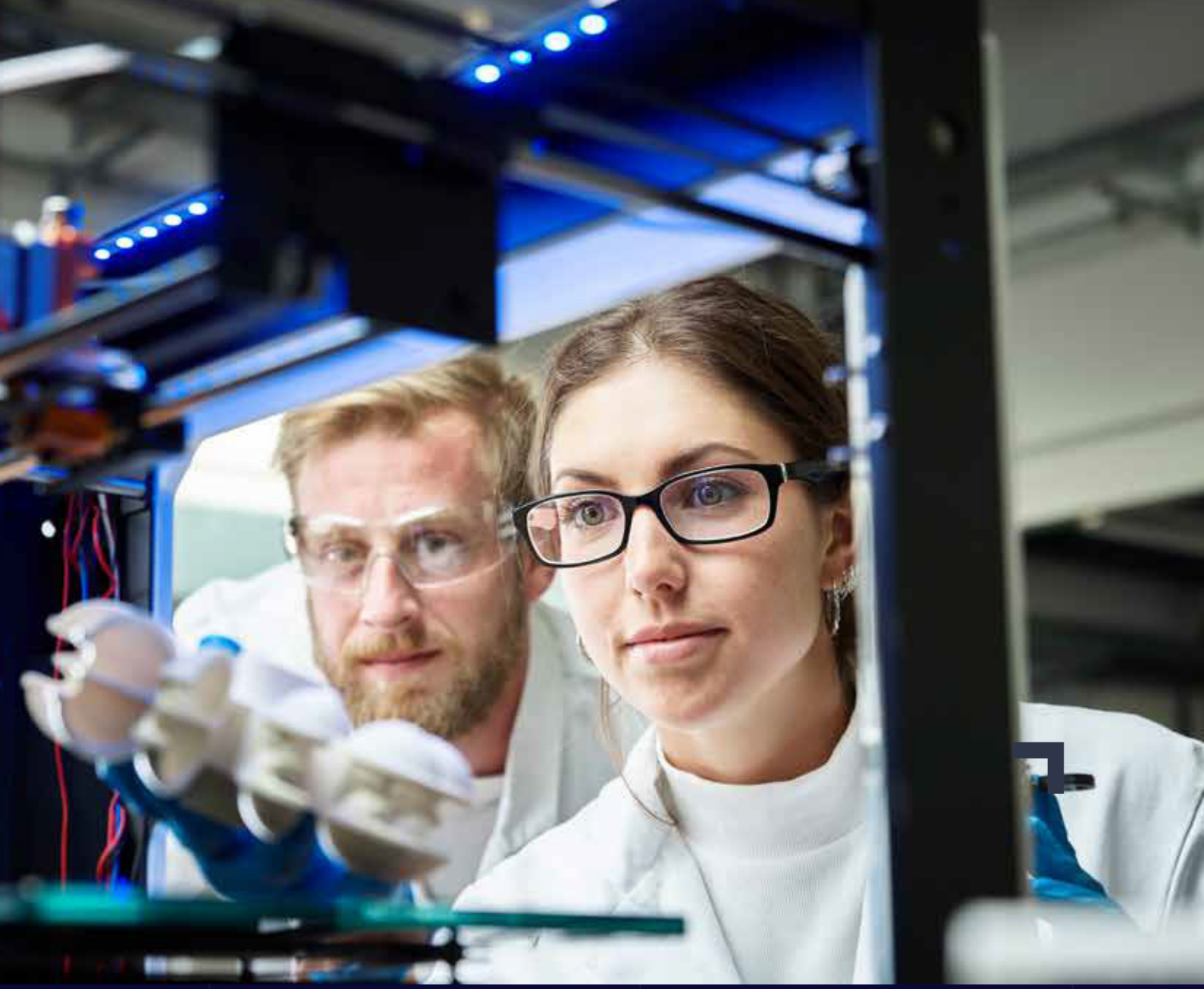

Whitepaper

# Dawn of Hi-Tech AI Factory

## Scaling Engineering Intelligence

Scale at Speed™

TECH mahindra

# Executive Summary

Artificial intelligence (AI) is not the next product; it's the next production line. Until organizations leverage an AI factory that ships trustworthy, governed, and measurable AI changes in an agile fashion, investments in isolated pilots will continue without the industrialization of intelligence in core business.

This whitepaper introduces the AI factory, an integrated and cohesive operating model designed to convert data into measurable business intelligence systematically. The AI factory is not a single tool or team, but a complete system encompassing the data supply chain, model lifecycle, governance, and human-in-the-loop processes. Its singular goal is to make the delivery of AI-powered capabilities repeatable, traceable, and reliable.

In this piece, we outline a clear, technology-neutral architecture for an AI factory, detailing the key components, critical roles, and a phased implementation roadmap. For senior leaders, this provides a blueprint to transform AI from a high-cost, high-risk endeavor into a disciplined, value-generating engine that drives continuous improvement and sustainable competitive advantage.

# Table of Contents

# Introduction

Industry analysts frequently report that a high percentage of AI projects fail to deliver on their intended business value, often getting stuck in what's often called "pilot purgatory." The challenge isn't a lack of brilliant data scientists or powerful algorithms; it's the absence of a disciplined, industrial-grade process for creating and managing intelligence.

It's time to stop treating AI as a series of science experiments and start treating it as a core engineering capability. The solution is to build an AI factory—an integrated and cohesive operating model that systematically converts data into valuable, production-grade intelligence. This isn't about a specific tool or team. It's a comprehensive system that provides the guardrails, traceability, and control necessary to create reliable AI pipelines that consistently drive measurable business impact.

# Defining AI Factory: From Bespoke Projects to Industrialized Production

The AI factory is an operating model that industrializes the process of creating and deploying AI capabilities. Our methodology moves organizations from inconsistent, artisanal projects to a standardized, production-line approach that predictably delivers value. The objective is to engineer reliable intelligence pipelines that translate ideas into measurable business impact while maintaining immutable guardrails for safety, ethics, and cost.

Our philosophy emphasizes a disciplined practice built on sound data management and responsible AI. The result is not simply "more models," but dependable capabilities embedded directly into your day-to-day operations—such as forecasting, anomaly detection, and recommendation—delivered with service levels you and your customers can trust.

# The Need for an AI Factory Now

The convergence of several key enablers makes the AI factory approach not just possible, but essential for competitive advantage today:

### Accessible Compute

On-demand cloud infrastructure can be right-sized for both training and inference.

### Mature Data Tooling

Advanced tools for data quality, cataloging, lineage, and privacy are now available at scale.

### Advanced Models

Libraries for classical ML, language, and vision models are widely available for diverse use cases.

### Robust MLOps Practices

Platform engineering principles allow for standardized builds, tests, and deployments.

These components alone don't create value. The AI factory connects them into a disciplined, product-like flow where outcomes are tested, versioned, and measured against your most important business targets.

## The TechM AI Factory

The AI factory we have built is a product line for intelligence. We are continually working to develop a method that provides standardized components, documented processes, and shared platforms, enabling high-performance (HiPo) teams to design, test, certify, and operate AI tools with the same reliability as we produce software services.

### Characteristics of the TechM AI factory

**Integrated data supply chains**

To deliver trustworthy inputs with lineage, quality checks, and access controls

1

**Model lifecycle**

To ensure trustworthiness from exploration to production with clear promotion gates and reproducibility.

2

**Runtime infrastructure**

Security-aware infrastructure that is elastic, observable, and policy-aware.

3

**Human-in-the-loop (HiTL)**

HiPo teams that act in HiTL mode for supervision, escalation, and improvement.

4

**Governance and Control**

Ensuring that legal, ethical, and security requirements are encoded directly into workflows.

5

**Economic discipline**

Strict control of costs and value is visible, attributable, and manageable.

6

We do not deliver "more models". We provide dependable capabilities that are embedded in day-to-day operations, including forecasting, detection, summarization, classification, recommendation, generation, and planning. All this, delivered with service levels that customers can trust.

## Our Reference Architecture – From Ingestion to Impact Creation

We have created a concise and modular reference architecture to help keep teams aligned while allowing for the choice of tools. It focuses on the following five elements:

- Data Supply Chain: Sourcing, controls, preparation, and lineage management.
- Model Lifecycle: Governed exploration, reproducible training, rigorous evaluation, and secure promotion.
- Orchestration and Runtime: Service management, policy enforcement, observability, and cost/capacity optimization.
- Human-in-the-Loop and Feedback: Review queues, escalation playbooks, and structured feedback capture for continuous learning.
- Governance and Risk: A central model registry, comprehensive audit trails, and codified policy packs.

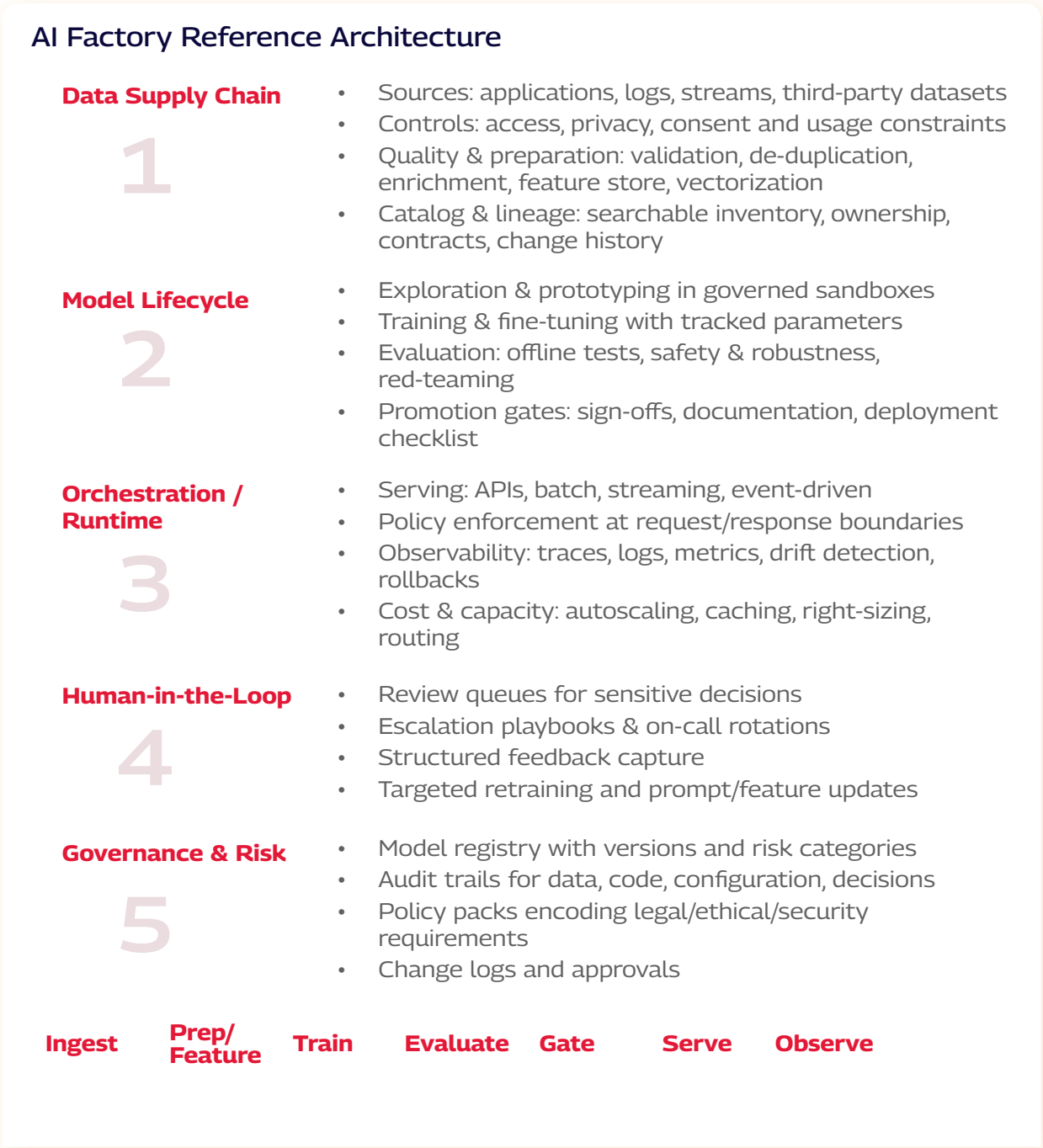


Figure 1: AI factory reference architecture from ingest to impact, with built-in governance and feedback.

# The Operating Model: People, Responsibilities, and Cadence

In our experience, clear ownership is the primary differentiator between successful production systems and perpetual science projects. Our AI factory implementation establishes well-defined roles and a disciplined operational cadence to ensure accountability and momentum.

### Key roles we help you establish

- Product Owners (POs): To champion business outcomes and define success.
- Data Stewards: To guarantee the integrity of the data supply chain.
- Applied Scientists / ML Engineers: To build the core intelligence features.
- Platform Engineers: To provide and maintain the shared infrastructure.
- Security, Legal and Compliance: To codify trust and safety requirements into the system.
- Site Reliability Engineers (SREs): To ensure operational excellence and manage SLAs.
- Domain Experts / Reviewers: To provide essential human judgment and oversight.

### Our recommended operational cadence:

- Weekly: Triage, experiment reviews, and change control.
- Monthly: Model health assessments, risk reviews, and value tracking.
- Quarterly: Roadmap refresh and portfolio reprioritization.

# Governance, Safety, and Assurance

Our AI factory is built on the principle that the responsible use of AI is a design requirement, not an afterthought. We've built governance into how the AI factory works:

### Key roles we help you establish

- Document intended use, limitations, as well as out-of-scope scenarios before shipping a use case.
- Use evaluation suites that test performance, reliability, robustness, and policy adherence.
- Apply defense-in-depth – input validation, output filtering, rate limiting, and sensitive-use routing.
- Maintain a changelog and audit trail for data, prompts, models, and configurations.
- Treat "Human in the Loop" (HITL) as part of the control system, with training and accountability.
- Keep all evidence ready and in place in accordance with applicable regulations and internal audit standards.

Our goal is to develop fit-for-purpose solutions. We are crafting systems that are reliable enough for their stated use and easy to maintain as conditions change.

# Measuring Success: Our Approach to AI Economics

For AI to be a core business capability, it must be judged with the same financial and operational rigor as any other production system. Our framework focuses on tangible metrics and disciplined financial operations (FinOps).

## Core metrics we track

- Outcome Metrics: Error reduction, cycle-time reduction, conversion uplift.
- Reliability Metrics: Latency, availability, SLO attainment, MTTR.
- Data Health Metrics: Freshness, completeness, quality, and lineage.
- Model Health Metrics: Drift, calibration, and safety scores.
- Unit Economics: Cost per request, workflow, or transaction.

## FinOps and Capacity Planning

- Tag cost to products, teams, or use cases.
- Prefer to select the lowest-cost backend that meets the quality thresholds for a given use case.
- Utilize caching, batch processing, and rate-aware routing to optimize spending as far as possible.
- Set budgets, alarms, and alerts, halt or degrade gracefully when limits are reached.

## Decision Principles: Build versus Buy

Our AI factory supports both internal builds and external services. We decide with simple, repeatable criteria:

- We aim to differentiate where it matters. We build when the capability is core to our advantage or requires domain-specific signals.
- We recommend buying and supporting it when the market provides sufficient quality, better security posture, or faster time-to-value.
- We try to avoid lock-in by isolating interfaces, using registries, and containerizing evaluation pipelines.
- We standardize contracts for data, models, and services. This makes swapping components a routine rather than a disruptive process.
- We maintain a small set of default patterns that are well-documented and well-supported.

# Our Phased Implementation Roadmap: Your Path to Scale

Our approach is designed to de-risk your investment, demonstrate value in weeks, not years, and allow the operating model to mature organically within your organization. The AI factory creates contextual and customer-specific implementation roadmaps for AI infusion.

## Phase 0 - Immediate

In this phase, we identify owners for products, data, platforms, and risk. We select two high-value, bounded use cases and establish a single backlog for them.

## Phase 1 - 30 Days

We create a stand-up on the minimal platform in the second phase. We develop a data catalog, model registry, CI/CD for models, policy enforcement at the edge, and establish basic observability. We define SLOs and error budgets for each use case being considered.

## Phase 2 - 90 Days

In this phase, we move the use cases to limited production. We instrument telemetry, capture feedback, and establish monthly model health and risk reviews. We document playbooks for on-call and rollbacks.

## Phase 3 -180 Days

This phase expands to a portfolio of 5–8 use cases on the same platform. We bring in cost allocation principles, standard evaluation suites, and quarterly roadmap reviews. We also publish a concise "factory handbook" to help new teams onboard quickly and smoothly.

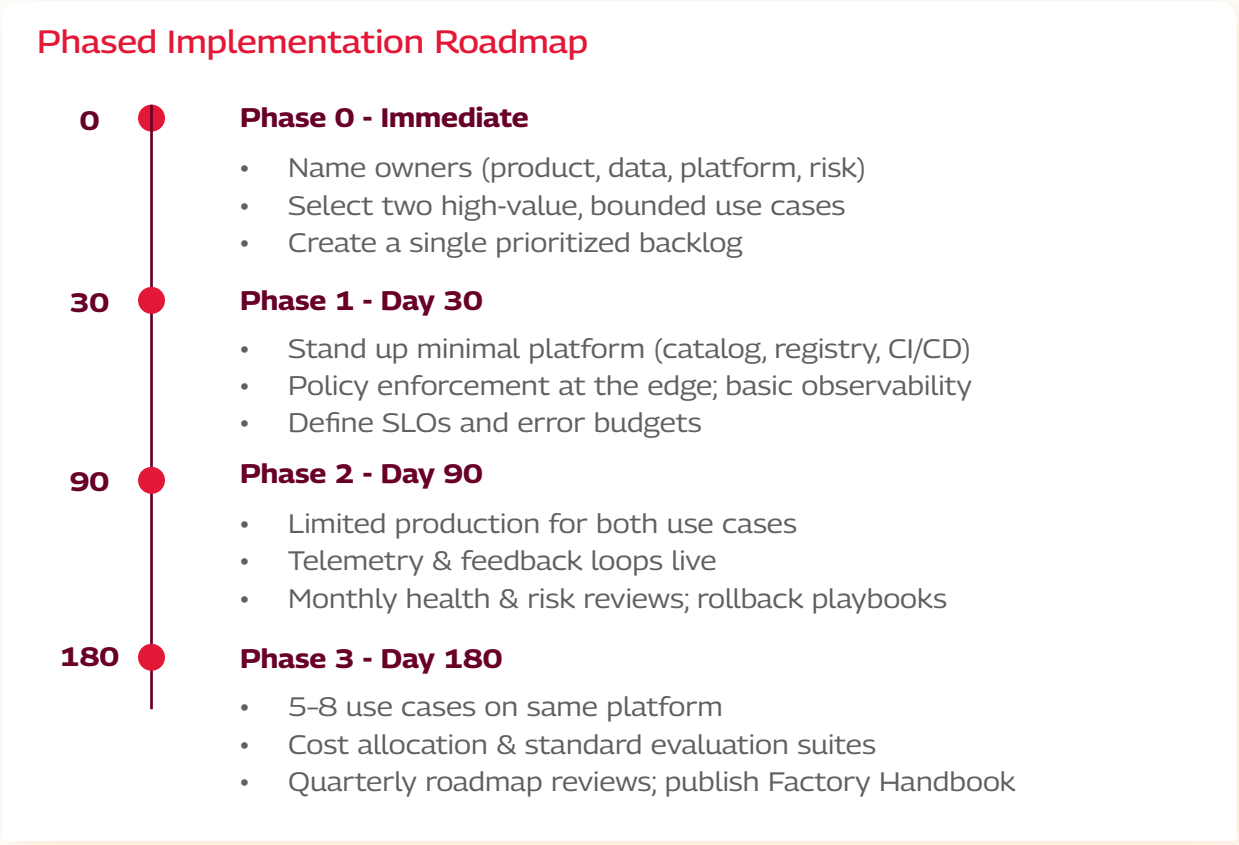


Figure 2: Phased rollout of the AI factory, from minimal platform to portfolio scale.

# Navigating Pitfalls: How Our Methodology Ensures Your Success

Through our work, we've seen firsthand the pitfalls that derail promising AI initiatives. Our AI factory methodology is explicitly designed to help you navigate and avoid these common failure modes.

- Tool-First Thinking - We recommend against accumulating platforms without a clear operating model. We recommend starting from outcomes, roles, and metrics.
- One-Off Projects - We don't advise doing pilots that never meet production standards. Instead, we run everything through the same promotion gates.
- Opaque Costs - We are very wary of surprise bills and unclear ROI. So as a principle, we attribute costs and set budgets early.
- Unmanaged Drift - We recognize that performance can decay silently. We continuously evaluate, set, and check alerts, and have stringent retraining schedules.
- Missing Guardrails - We don't want to leave scope for policy violations or security gaps. We enforce controls at interfaces, log decisions, and review them regularly.
- Over-centralization - We abhor bottlenecks and slow iterations. We aim for central guardrails with federated product ownership and paved-road tooling for each customer.

# Why This Matters in Hi-Tech, Media, and Entertainment (TME)

We implemented this operating model for the AI factory for TME enterprises. It is repeatable across verticals, but some nuances are specific to TME.

## Hi-tech characteristics

- Rapid release cycles and complex product portfolios in the hi-tech sector favor standardized evaluation, promotion gates, and telemetry.
- We have found that hardware-software convergence benefits from rigorous data contracts, simulation pipelines, and feedback-driven refinement.
- We have benefited from ecosystem partnerships and channels. These require policy-aware interfaces and auditable behavior across boundaries.
- We have evidence that customer support, quality, and supply chain workflows gain from retrieval, classification, forecasting, and recommendation at scale.

## Media and entertainment characteristics

- From ingestion to editing, rights management to packaging and distribution, the stages of the content supply chain align neatly with the AI factory stages.
- There are benefits from personalization and discovery through transparent evaluation, safety reviews for sensitive content, and runtime policy controls. Safe-by-design principles embedded in the AI factory work well for media and entertainment.
- Live operations, such as streams, sports, and events, demand predictable latency, rollback strategies, and clear escalation paths. This is the strength of the AI factory design.
- Advertising, measurement, and rights require auditable data flows and explainable decisions. The governance policies are well-aligned with this request.

In both sectors, the AI factory improves speed without losing control. It enables reproducibility and governance simultaneously with experimentation. This balance is difficult to improvise. It is straightforward to engineer when the AI factory is explicit, as in our case.

## Best Practices at Our AI Factory

We have learnt that when the AI factory is healthy, teams share patterns instead of reinventing them. Our use cases ship with consistent documentation. Our promotion gates are respected. Nearly all incidents lead to improvements in playbooks and controls that become practice. Our costs are predictable and defensible. Evaluation is part of our day-to-day work, not a special event. We can onboard new teams quickly because the path is paved and the obligations are clear.

The most important property of our factory is the learning rate. The AI factory improves because telemetry, feedback, and reviews are built into the loop of how we operate. That is what turns individual projects into a durable capability.

## Your Partner in Building a Durable AI Capability

Our AI factory proves that there is a practical way to create dependable and measurable intelligence in production. We align people, data, models, and platforms around repeatable outcomes. We keep governance in the foreground. We make costs visible. We create a shared language for how ideas evolve from exploration to impact.

We always recommend starting small but designing for scale. We name clear owners. We select a few valuable use cases and refer to them as lighthouses. We build the minimal shared platform and enforce promotion gates. We measure outcomes and costs. We review regularly. We expand only when the path is smooth.

# Authors

**Dr. Anshu Premchand**
VP & Group Function Head
Multi Cloud & Digital Services
Technology, Media & Entertainment Unit
Tech Mahindra

Dr. Anshu Premchand is a persuasive thought leader with 25+ years of experience driving AI, cloud, and digital transformation at Fortune 1000 enterprises. As Vice President & Group Function Head for AI, Cloud & Digital Services in Tech Mahindra's Hi-Tech, Media & Entertainment unit, she leads multi-cloud strategy, AI-driven innovation, and large-scale transformation programs.

She has spearheaded enterprise-wide modernization initiatives across data, applications, infrastructure, and security, leveraging AI and automation to accelerate digital adoption in the past. She has successfully designed and executed multi-million-dollar cloud transformation programs, delivering tangible business impact across BFSI, Manufacturing, Retail, and Utilities.

Anshu holds a Ph.D. in Computer Science and has an extensive publication record. She is passionate about IT simplification, predictive operations, and the future of AI-driven enterprises."

**Sukumar Shanmugam**

Delivery Head
Tech Mahindra

Sukumar Shanmugam is a Proven Technocrat and Hi-Tech Delivery Head with 20+ years of experience in global delivery and operations. He currently manages a strategic portfolio of Hi-Tech accounts for Tech Mahindra, specializing in building client relationships and scaling high-performance teams. Sukumar possesses deep technical expertise in AI, Cloud, Data Platforms, and Full Stack Engineering. His career spans multiple sectors (Telecom, Banking, Capital Markets, Insurance, Hi-Tech), where he consistently won and successfully delivered large-scale technology programs.

# About Tech Mahindra

Tech Mahindra (NSE: TECHM) offers technology consulting and digital solutions to global enterprises across industries, enabling transformative scale at unparalleled speed. With 152,000+ professionals across 90+ countries helping 1100+ clients, Tech Mahindra provides a full spectrum of services including consulting, information technology, enterprise applications, business process services, engineering services, network services, customer experience & design, AI & analytics, and cloud & infrastructure services. It is the first Indian company in the world to have been awarded the Sustainable Markets Initiative's Terra Carta Seal, which recognizes global companies that are actively leading the charge to create a climate and nature-positive future. Tech Mahindra is part of the Mahindra Group, founded in 1945, one of the largest and most admired multinational federation of companies. For more information on how TechM can partner with you to meet your scale at speed imperatives, please visit https://www.techmahindra.com/

www.techmahindra.com
www.linkedin.com/company/tech-mahindra
www.x.com/tech_mahindra
TMEmarketing@TechMahindra.com