ecosystm.

Data Readiness for AI: Building the Foundation for Enterprise-Scale AI

JUNE 2026

Table of Contents

Introduction

Value creation is becoming the real test of enterprise AI – and it's starting to reshape how organisations think about strategy itself.

Across Asia Pacific, AI is not just confined to innovation teams or isolated pilots. Market leaders are building it into core business strategy, influencing decisions on customer engagement, operations, risk, and productivity. As this happens, AI stops being just a technology agenda and starts driving changes in how organisations are actually structured and run.

That shift matters. Because when AI becomes part of business strategy, it inevitably pulls in operating model questions: how decisions are made, how work flows across functions, and how accountability is distributed between business and technology teams.

This is where many organisations are starting to feel friction. They can define the strategy, and they can fund the use cases, but they struggle to make the organisation move in a way that allows AI to deliver consistent value.

The biggest challenge that most organisations face is in the data layer that underpins these operating models: how data is accessed across teams, how consistently it is defined and governed, and whether it can move through the organisation in a way that supports decisions and processes without repeated manual intervention or reconciliation.

66%
of organisations are experimenting with or deploying AI use cases

ONLY 10%
are seeing measurable business value from those investments

JUST 6%
have data pipelines capable of supporting AI workloads in a consistent, scalable way

Source: Ecosystm, 2026

What this suggests is fairly straightforward. AI success is no longer just about models or tools. It depends on whether organisations can align three things at once: strategy, operating model, and data. And in most cases, it is the data layer that determines whether that alignment actually holds.

Why Enterprise Data is Not Built for AI

Most enterprise data environments were not originally designed for AI workloads. They were built to support transactional systems, structured reporting, and human-driven analysis.

These systems perform well when data flows are predictable, queries are defined, and usage patterns are stable – in other words, when outcomes are largely deterministic.

AI introduces a different operating reality. Many AI systems are non-deterministic: the same input can produce different outputs depending on context, model behaviour, and the data it draws from. That fundamentally changes what "good data" needs to look like when solutions are scaled.

AI systems move away from periodic data extraction towards continuous access to high-quality data across multiple domains. They go beyond structured datasets alone, requiring a mix of structured, unstructured, and semi-structured data. And instead of relying on system-specific views, they depend on cross-system context to produce outputs that are consistent, relevant, and explainable.

This shift exposes gaps that were previously manageable in traditional analytics environments but become material barriers in AI-driven workflows.

As organisations expand their AI footprint, three structural realities emerge:

- Data is distributed across legacy systems, modern cloud platforms, and third-party environments
- Integration between these systems is inconsistent and often designed for specific use cases rather than enterprise-wide reuse
- Ownership of data is fragmented across functions, making consistency and governance harder to enforce at scale

This explains why organisations can still demonstrate AI success in isolated environments. In constrained, well-bounded use cases, data dependencies are limited, integration points are defined upfront, and the system context is narrow enough to mask underlying fragmentation. The challenge emerges when those same systems need to operate across domains, where consistency, traceability, and shared context become mandatory rather than optional.

Asia Pacific Data Reality

Across Asia Pacific, while most enterprises are actively investing in AI, the ability to replicate outcomes consistently across the organisation remains uneven.

The underlying issue is the variability in enterprise data environments once AI is required to operate across systems, functions, and governance boundaries.

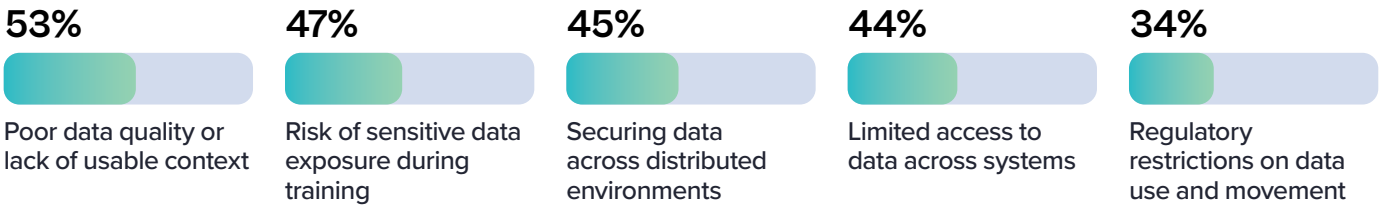
Structural Constraints in AI Environments

As organisations attempt to operationalise AI, a consistent set of challenges emerges. These are not isolated technical issues. They reflect how enterprise data environments behave under real-world operational load, where multiple systems, governance requirements, and business processes interact simultaneously.

Leaders consistently say that the challenge is not just enabling access to data but also maintaining reliability, consistency, and control across that access in live environments.

FIGURE 1

Key Operational Barriers to AI Adoption



Source: Ecosystm, 2026

These challenges show up at different stages of the AI lifecycle, but they are tightly coupled in how they shape delivery.

Data quality issues are most visible during use case preparation and onboarding. The effort required to source, clean, and reconcile data increases when datasets sit across multiple systems with inconsistent definitions, ownership, and refresh cycles. In many organisations, these stop being a one-off exercise and become something that repeats each time a use case is extended or a new data source is introduced.

Access determines what data AI systems can actually draw on. When relevant data is spread across business units, platforms, or external environments, availability becomes uneven by design rather than by exception. This limits both the breadth of context available to models and the reliability of outputs in workflows that depend on end-to-end visibility.

Security requirements add operational overhead as data moves across cloud, on-premises, and third-party environments. Controls are often implemented as additional validation steps layered across systems that were never designed to operate under a unified governance model. This introduces coordination overhead and increases latency in production workflows.

Regulatory constraints shape system design from the outset. Data residency rules, usage permissions, and cross-border restrictions determine how data is segmented, where it can be processed, and what can be combined. In practice, compliance requirements often introduce additional processing and movement steps that sit on top of the core system design.

Individually, these challenges are usually addressed through targeted fixes. However, when multiple solutions are scaled, they accumulate rather than remain isolated. The combined effect is non-linear: effort compounds across preparation, access, governance, and compliance layers, making it progressively harder to move AI from initial deployment into sustained, repeatable production use because the surrounding data environment introduces persistent structural friction.

Enterprise Data Maturity is Uneven Across Asia Pacific

Data maturity varies widely, with most organisations still operating in partially integrated or fragmented landscapes.

FIGURE 2

State of Enterprise Data Foundations

30%
Fragmented data landscape with manual access

22%
Centralised storage with limited real-time access

17%
Connected core systems through an integration layer

25%
Defined data model with shared business context

6%
Operationalised data pipelines with automation and monitoring

Source: Ecosystm, 2026

More than half of organisations are still operating in environments where data is fragmented or only partially integrated. These setups are adequate for reporting, dashboards, and early-stage AI experiments, but are not designed for workloads that depend on consistent, real-time, cross-domain access.

Even where integration layers exist, a common limitation remains: systems may be connected but not aligned. Differences in definitions, ownership, and governance mean that data can be technically available but still inconsistent in meaning and context when used by AI systems.

At the other end of the spectrum, only a small proportion of organisations have fully operationalised data pipelines with embedded automation, monitoring, and governance. These are currently the only environments structurally capable of supporting AI reliably across multiple functions and domains.

Variability Within Enterprises is a Reality

Even within individual organisations, data maturity is rarely consistent across functions. Some lines of business operate on modern, well-integrated data platforms, while others remain dependent on legacy systems, manual extracts, or siloed data stores. As AI use cases expand across business functions, they inevitably span both mature and immature environments. This creates uneven performance across use cases, even when the underlying models are identical. This internal variability is one of the most persistent issues hampering scaling AI across the enterprise.

Data Readiness & the Economics of AI

The next question is how the challenges translate into economic terms. Data readiness is not only a technical dependency; it determines how expensive AI becomes to run and whether it can scale beyond a handful of use cases.

In most organisations are willing to experiment most organisations, a large share of effort sits in getting data into a usable state across systems – aligning definitions, resolving inconsistencies, and building the pipelines needed for deployment. This work is rarely captured cleanly in AI budgets, but it drives both delivery timelines and ongoing costs.

Where data environments are fragmented, this effort is repeatedly recreated. Each new use case effectively restarts parts of the same work: extracting data from multiple systems, reconciling differences in meaning, and rebuilding pipelines that are not designed to be reused. The result is duplication across teams and a steady accumulation of complexity that sits underneath the AI portfolio.

In more mature environments, the pattern changes. Shared data models, reusable pipelines, and consistent governance mean new use cases can plug into existing foundations rather than rebuild them. That shifts effort away from setup and towards extension. It also means AI systems are working from more consistent inputs across the organisation, which improves reliability without requiring additional model-level intervention.

This is where the economic gap becomes visible. Where reuse is possible, AI capability compounds across functions. Where it is not, value remains trapped in individual deployments, with each one carrying its own cost base.

Agentic AI Raises the Stakes on Data Readiness

The move towards agentic AI changes the nature of the problem again.

Until now, most enterprise AI has been about generating outputs for human interpretation. The system analyses data, produces a recommendation or prediction, and a person decides what happens next.

Agentic systems remove that separation. They are designed to take actions across enterprise systems, triggering workflows, updating records, initiating transactions, and coordinating steps across multiple applications without continuous human intervention.

When AI is used for recommendations, data issues tend to surface as accuracy or quality problems in outputs. When AI is used for execution, those same issues translate into operational consequences. A data inconsistency becomes a workflow decision, a transaction error, or a downstream system impact. This significantly raises the dependency on data consistency, governance, and real-time reliability.

It also increases the exposure created by fragmentation: what was previously contained within a single use case now propagates across systems and processes. This moves data readiness moves a performance consideration to an operational risk boundary.

This is reflected in the barriers organisations are already identifying as they explore agentic AI.

FIGURE 3

Top Barriers to Adopting Agentic AI

53% Regulatory compliance across autonomous decisions

49% Ensuring data readiness and integrity

43% Maintaining transparency and explainability

37% Defining appropriate levels of human oversight

34% Ensuring security and preventing misuse

Source: Ecosystm, 2026

These map directly back to the underlying data environment. Each barrier is a reflection of how data is defined, accessed, governed, and trusted across systems. Agentic AI does not introduce an entirely new category of challenges; it amplifies existing ones. In environments where data is already fragmented or inconsistently governed, the shift from “AI that advises” to “AI that acts” removes the remaining buffer between data issues and business outcomes.

Building Data Readiness for AI

Most organisations no longer struggle to identify AI opportunities. The challenge is creating an environment where those opportunities can be deployed, operated, and scaled consistently.

As AI becomes embedded in day-to-day operations, the underlying data environment becomes a critical determinant of performance. Data readiness reflects the combined maturity of architecture, execution capability, governance, and data management practices that allow AI to operate reliably across the enterprise.

Architecture: The Structural Foundation

Traditional enterprise architecture was built around integration layers, data warehouses, and APIs. It was assumed that data would be consumed in structured, predictable ways, primarily through human-driven reporting, dashboards, and application queries.

AI breaks this assumption. Data is no longer simply retrieved; it is continuously interpreted, combined across sources, and used as input into systems that generate decisions or trigger actions. This changes what "usable architecture" means.

At a minimum, AI-ready architecture needs to move beyond system connectivity and address how data behaves across environments.

FIGURE 4

The Architectural Foundations of AI-Ready Data

Source: Ecosystm, 2026

Distributed as the default operating state

Enterprise data is spread across cloud platforms, on-premises systems, SaaS applications, and external ecosystems. Rather than trying to centralise everything, organisations need architectures that can work effectively across distributed environments, enabling seamless access to data without constant replication or manual reconciliation.

Consistency of context across systems

Connecting systems is only part of the challenge. AI depends on data having the same meaning wherever it is used. Shared definitions, aligned metadata, and common business context are essential to ensure models are working from a consistent view of the organisation rather than conflicting versions of the same information.

Governance embedded in data movement

Governance cannot be treated as a separate control function applied after the fact. Access controls, lineage tracking, retention policies, and usage rules need to be embedded directly into how data moves through the organisation, ensuring compliance and accountability are maintained continuously rather than checked periodically.

These shifts have practical implications for how data environments are designed. They influence where data is accessed, how AI systems interact with it, and how consistency is maintained across distributed environments without forcing everything into a single platform.

Orchestration: How Work Actually Moves Across Systems

Once data is accessible and governed across environments, the next issue is more operational: most enterprise processes do not live in a single system anymore.

A single workflow might start in a CRM, pull context from a data platform, trigger a decision in an AI service, and complete in an ERP. These steps are loosely connected, owned by different teams, and built on different assumptions about timing, structure, and control.

This is where things start to break when AI is scaled, not because systems are disconnected, but because there is no consistent way to manage how work moves between them once AI is involved.

Orchestration is the response to that gap. It is about making cross-system execution predictable enough to operate reliably across environments, without forcing everything into a single platform or workflow engine.

This means:

- **Work does not follow a single path.** The same process may take different routes depending on data availability, system response, or policy guidelines.
- **Execution happens in multiple places.** Different steps run where the data or capability exists, rather than being pulled into one central environment.
- **Responsibility is distributed.** No single system "owns" the full process, which makes traceability and control a design requirement, not an assumption.

To make this manageable, organisations are converging on a few practical design patterns:

DECISIONS BASED ON CONTEXT, NOT HARD-CODED FLOWS

Instead of fixed integration chains, systems decide where a task should go based on available data, rules, and conditions.

SEPARATION OF TRACKING FROM EXECUTION

The system that monitors a process is not the same as the systems that perform the steps, which allows components to change without breaking the workflow.

A SINGLE VIEW OF PROGRESS ACROSS SYSTEMS

Even though execution is distributed, organisations still need to see what is happening across all steps without logging into multiple tools.

The goal is not to introduce another orchestration layer for its own sake. It is to avoid the current reality where cross-system processes only work reliably when everything is manually aligned in advance.

Execution Capability in AI Environments

Even when architecture is sound and orchestration is in place, AI systems often behave differently once they are operating in live environments. Execution capability is what determines whether AI remains stable after deployment. It reflects how reliably data, governance, and system interactions hold together under real operating conditions.

FIGURE 5

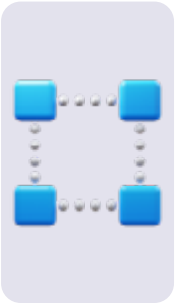
Execution Reality Layers in Enterprise AI Environments

Data Flow Reality	Runtime Control	Execution Constraints
01. Pipeline integrity & real-time readiness	01. Governance as runtime behaviour	01. Sovereignty & compliance requirements
02. Coherence under operational load	02. Observability as operational control	02. Privacy-preserving data use

Source: Ecosystm, 2026

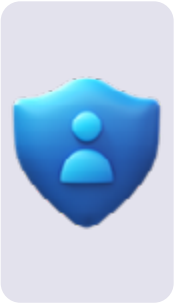
Pipeline integrity and real-time readiness

AI systems depend on data that reflects the current state of the business, not delayed extracts or periodic refresh cycles. Many enterprise pipelines were designed for reporting and analytics, where timing gaps are acceptable and correctness is assessed after the fact. In AI-driven environments, that tolerance is significantly reduced. Even small delays, gaps, or inconsistencies in data movement can lead to unstable or non-repeatable behaviour in downstream decisions. This is not about pipeline speed alone. It is about whether data flows are dependable enough to support continuous decision-making without manual correction or reconciliation.



Coherence under real-world load

Distributed data environments behave differently under operating conditions than they do in design assumptions. Differences in definitions, metadata, and interpretation often only become visible when multiple systems are combined in active workflows. Individually, systems may appear consistent; together, they diverge. Sustained AI performance depends on whether these inconsistencies remain contained or begin to affect outcomes across functions.



Governance as part of runtime behaviour

Governance cannot sit outside execution. Access rules, usage constraints, and audit requirements must be enforced as data moves, not after it has been processed. Where governance is applied as a separate control function, gaps emerge between intended policy and actual system behaviour. These gaps widen as the number of systems and interactions increases. Governance must therefore shift from a compliance layer to a runtime property of the system itself.

Observability as operational control

Once AI operates across distributed systems, visibility becomes the only reliable mechanism for control. Observability now extends beyond infrastructure health into data movement, pipeline behaviour, and system-level interactions. It is what allows organisations to detect failure modes that only appear when multiple systems interact under load. It is not diagnostic support. It is part of the operating model.

Sovereignty and compliance as execution guardrails

Regulatory requirements are shaping how workflows are executed, not just how data is stored. Data residency rules, cross-border restrictions, and usage limitations influence routing, processing location, and system interaction patterns. In many cases, compliance requirements introduce additional logic directly into execution paths. This makes regulation a structural input into how systems run, not an external factor applied afterwards.

Privacy-preserving data use

As AI expands into domains involving sensitive or regulated data, direct exposure of raw datasets becomes increasingly restricted. Techniques such as synthetic data, federated learning, and privacy-preserving inference allow systems to generate and use insights without moving or exposing underlying data. This removes the limitations of data availability.

Data Governance for AI

While enterprise AI governance is often framed in terms of ethics, model behaviour, and accountability, data governance is what determines whether data is reliable, controlled, and safe to use across distributed environments. As AI becomes embedded in core business systems, governance is shifting away from documentation-heavy oversight toward mechanisms that are built into system behaviour.

For governance to remain effective in live environments, it must operate consistently across both architecture and execution. Gaps emerge when these layers are designed independently and do not reinforce each other. The architecture layer defines how data is structured, exposed, and made available across distributed, multi-platform environments. The execution layer defines how data is accessed, constrained, and routed as it moves through operational systems. When these layers are not aligned, systems behave inconsistently across environments. This is rarely due to model failure. More often, it reflects differences in how data is defined, interpreted, or enforced across systems.

Data governance is evolving along two practical dimensions that determine how control is implemented in real-world AI environments.

01. Shift to Embedded Runtime Controls

Rather than relying on periodic audits or post-event policy enforcement, governance needs to move closer to the data itself, operating continuously at the dataset, pipeline, and event level. To achieve this, organisations should embed governance directly into the data layer across four areas:

- **Point-of-use access control**
  Security and access rules are enforced dynamically at the moment data is used, rather than only at system entry points.
- **Continuous lineage capture**
  Data origin, transformation, and movement are tracked automatically and continuously as information flows across distributed environments.
- **Active processing rules**
  Policy, residency, and retention requirements are embedded into execution paths so they shape live processing behaviour rather than remaining static rules applied at storage level.
- **Real-time signal generation**
  Governance and compliance signals are generated during data movement, replacing retrospective audit reconstruction with continuous visibility.

02. Minimum Viable Governance

Because data environments expand quickly, governance cannot scale in a linear way with system complexity. Trying to apply the same level of control across all data creates operational friction without improving risk outcomes. Organisations need a differentiated model where control is concentrated where exposure, risk, and business impact are highest.

FIGURE 6

Priority Areas for Minimum Viable Data Governance

Sensitive and regulated datasets	Highest compliance exposure and strict privacy requirements
High-impact AI systems	Models influencing critical operational or financial decisions
Cross-border data flows	Multi-jurisdiction pipelines subject to differing regulatory regimes
Third-party integrations	External models and systems operating beyond organisational control
Autonomous data pipelines	Data streams feeding autonomous or semi-autonomous workflows

Source: Ecosystm, 2026

Where Governance Meets Sensitive Data

A key challenge for enterprise data governance emerges where AI intersects with sensitive data. As organisations move from general-purpose datasets to proprietary customer, operational, and regulated information, the issue is not availability, but whether it can be used within privacy, compliance, and control boundaries. Governance sets the rules; execution determines how consistently those rules are applied in practice. This shifts the challenge from access to system design.

When AI systems operate on sensitive or regulated data, uniform access models begin to break down. Risks of exposure can arise during training, fine-tuning, and inference, including unintended leakage or inference of sensitive attributes. Relevant datasets are also distributed across jurisdictions and platforms, making consistent access difficult to enforce. Differing data protection requirements further prevent a single governance model from applying across environments. At the same time, tighter compliance controls can limit what data is available to models, affecting the range and reliability of outputs.

Rather than addressing this by widening access, organisations are embedding privacy and governance requirements directly into system design and execution. This has led to three dominant production approaches.

Synthetic data is being used where real data cannot be exposed. Instead of relying on actual customer, transactional, or operational records, organisations generate statistically representative datasets that preserve structural properties without exposing underlying entities. In financial services, this allows fraud and anti-money laundering models to be trained and tested without sharing sensitive customer data across institutions. In healthcare, synthetic patient and imaging datasets support research where access to clinical records is tightly restricted. Retailers are using similar techniques to train computer vision systems without relying on live store footage or identifiable customer data, while energy organisations simulate demand and grid behaviour to test scenarios without exposing critical infrastructure details.

Federated learning takes a different approach by removing the need to centralise data altogether. Instead of aggregating datasets into a shared environment, models are deployed to where the data already resides. Training happens locally within each environment, and only aggregated model updates are shared back to a central system. This approach is particularly relevant in regulated, multi-institution contexts. For example, in cross-bank fraud detection, institutions cannot exchange raw transaction data, but they can still contribute to a shared model that learns from distributed patterns while keeping data within its original boundary.

Differential privacy addresses a different but related risk: the tendency of large models to inadvertently memorise details from their training data. By introducing controlled statistical noise during training, it limits the influence that any single record can have on model outputs. The result is a measurable mechanism for balancing model utility with privacy protection, particularly in environments where models are trained on proprietary or regulated datasets.

These approaches are not experimental enhancements. They are becoming core components of how enterprise AI architectures are designed. Most organisations are not choosing one method in isolation but combining them to create layered protection across different stages of the AI lifecycle.

Implications

FOR ENTERPRISE LEADERSHIP

AI readiness is not determined driven by incremental improvements in data platforms or infrastructure. It depends on how ownership is distributed across IT and data, and whether the organisation can operate as a coherent system across distributed environments. Most friction now comes from fragmented ownership of tightly coupled decisions.

Treat AI readiness as a shared IT–data accountability model

AI outcomes are shaped as much by infrastructure as by data design. Decisions on cloud architecture, integration patterns, deployment models, and platform design directly determine data access, latency, and consistency. AI readiness cannot sit within a single function. It requires shared accountability across infrastructure, data, security, and application teams, with ownership defined through outcomes, not hand-offs.

Build governance into system design, not post-deployment review

Governance is moving into system design. Choices around access, usage, retention, and auditability now shape architecture and pipelines from the outset. This shifts governance from a control function to a design feature, where risk is engineered during build rather than assessed after deployment. Coordination between data governance, security, and platform engineering is structural.

Design for coordination across distributed data environments

Most enterprises will not centralise data. It remains distributed across cloud, legacy systems, SaaS platforms, and external ecosystems. The challenge is coordination, not consolidation. This requires consistent definition, access, and interpretation across environments not designed to interoperate. Without it, AI is constrained to areas where data is already aligned. The focus shifts from a single source of truth to consistent behaviour across sources.

Manage AI as an end-to-end system, not a set of components

AI performance is determined by how data, infrastructure, and applications interact under real operating conditions, not by improvements in isolated components. Latency, resilience, compliance, and interoperability operate as linked dependencies rather than separate levers. AI outcomes cannot be owned by individual teams. They reflect end-to-end system design in practice, where progress depends more on removing friction across the system than on adding capability.

Treat trust as a cross-stack design responsibility

Trust cannot be confined to governance or risk teams. It is shaped across infrastructure, data pipelines, application design, and model deployment. Issues such as bias, leakage, and lack of transparency emerge at system boundaries, not within isolated layers. It requires shared responsibility across IT, data, security, and product teams. It cannot be validated after deployment; it must be designed in.

FOR TECHNOLOGY PROVIDERS

Enterprise AI adoption is increasingly constrained by data readiness challenges that most organisations cannot resolve independently. The challenge lies in the ability to integrate, govern, and operationalise data across distributed environments in a consistent way.

Providers that move closer to the data layer – enabling integration across fragmented systems, enforcing governance in motion, supporting real-time access, and maintaining coherence across distributed architectures – are becoming central to enterprise AI strategies. Their role is shifting from adding capabilities on top of existing stacks to shaping how data is actually usable across them.

Expectations are evolving:

- Observability must extend across both data and AI workflows, not remain confined to infrastructure monitoring
- Interoperability across cloud, SaaS, and legacy environments is now a baseline requirement, not a differentiator
- Privacy-preserving capabilities are becoming mandatory in regulated industries rather than optional enhancements
- Abstraction of complexity is a primary driver of adoption, particularly where data fragmentation cannot be eliminated quickly

Technology companies that reduce structural friction in how data is accessed, governed, and used will be better positioned than those that layer AI capabilities over fragmented and inconsistently governed environments.