LET'S PRACTICE

THE AI-READINESS PLAYBOOK

Activate Measurable Foundations for AI in ~~Months~~ Weeks!

By Authors of the DATA PRODUCT PLAYBOOK
In collaboration with Data Leaders, Strategists, & Consultants

FOR ENTERPRISE AI ASPIRANTS

Modern Data 101 Community

TABLE OF CONTENTS

Modern Data 101 Community

Your roadmap to building production-ready AI data products in six weeks.

FOUNDATIONS

BUILD YOUR FOUNDATION

THE 6-WEEK PLAYBOOK

WEEK 1

WEEK 2

WEEK 3

WEEK 4

WEEKS 5 & 6

ADVANCED

Join the Modern Data 101 Community

20,000+ members • 130+ countries • 150+ contributors
400+ resources on data and AI

Learn. Share. Grow. Together.

STARTING WITH A REALITY CHECK

AI is the most hyped enterprise capability of the decade.
It is also the most under-delivered. And the gap between ambition and outcome is singularly a data problem.

Generative AI, large language models, real-time recommendation engines, predictive analytics. Enterprises are racing to integrate these capabilities at a pace that would be impressive if the results weren't so consistently disappointing. Gartner estimates that organisations will abandon 60% of AI projects through 2026 due to inadequate data infrastructure and unclear business value.

The irony is that the organisations spending the most on AI tooling like model APIs, vector databases, fine-tuning infrastructure, or LLMOps platforms, are often the least equipped to use them effectively, because they haven't solved the more fundamental challenge: **getting their data into a state that AI can reliably consume.**

85%	**60%**	**3-5 x**
enterprise AI projects fail to reach production of	*of data science time is spent cleaning & finding data*	*longer to ship AI features without data product infrastructure*

Modern Data 101 Community

In the rush to deploy AI, we've skipped a step. That step is productising data for AI consumers, and it is no longer optional. Fortunately or unfortunately.

This playbook is the sequel to **The Data Product Playbook**, which gave you the 6-week framework for activating data products in your organisation. If you have not read Part I, we recommend starting there. This playbook builds directly on that foundation and takes you through the next horizon: turning your data product ecosystem into an enterprise AI capability.

WHY ENTERPRISE AI ISN'T TAKING OFF

Modern Data 101 Community

If you ran an analysis of failed AI projects across enterprises, you would find the same five patterns appearing with almost metronomic regularity. None of them is about the model.

01

Data Inconsistency

Training and serving pipelines pull from different data sources or at different times, creating undetected model drift from day one.

02

Feature Entropy

Features are built once, never catalogued, and rebuilt repeatedly by different teams: duplicated effort, inconsistent logic, no single source of truth.

03

Governance Gaps

No lineage, no data contracts, no quality SLOs. When a model makes a bad decision, there is no audit trail to understand why.

04

Bottlenecked Access

Every AI team needs to wait for the central data team to get the data they need. Experimentation dies in the ticket queue.

05

No Feedback Loop

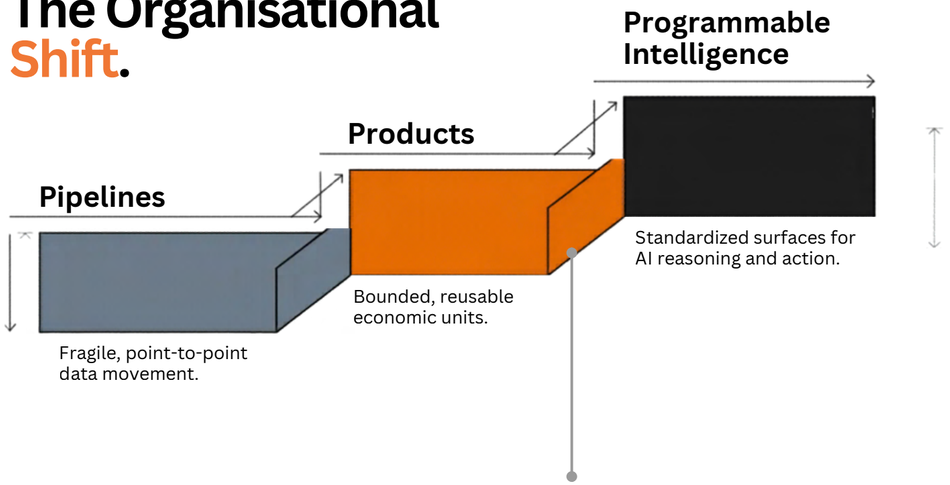
Model performance degrades in production with no systematic signal back to the data layer. Data never improves from real-world outcomes.

!!

The Common Thread

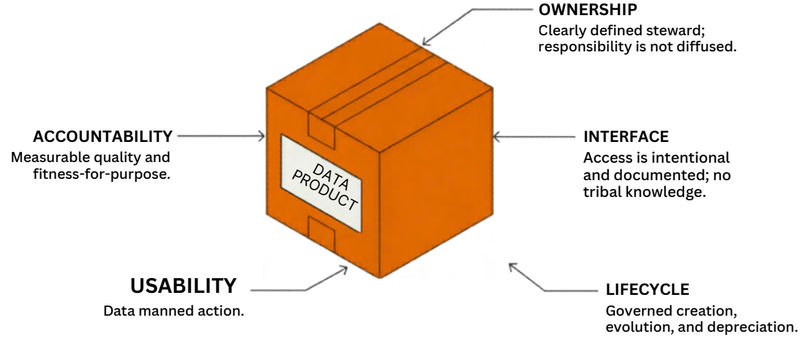
Each of these patterns is a data infrastructure problem and **NOT** a model problem. They are all solved by productising your data layer.

The Organisational Shift.



The Economic Plane: Data Products defined by Purpose

A Data Product is not a pipeline. It is a bounded, reusable economic unit of data, unified around a specific outcome.



AI IS A USER. DATA IS THE PRODUCT.

Modern Data 101 Community

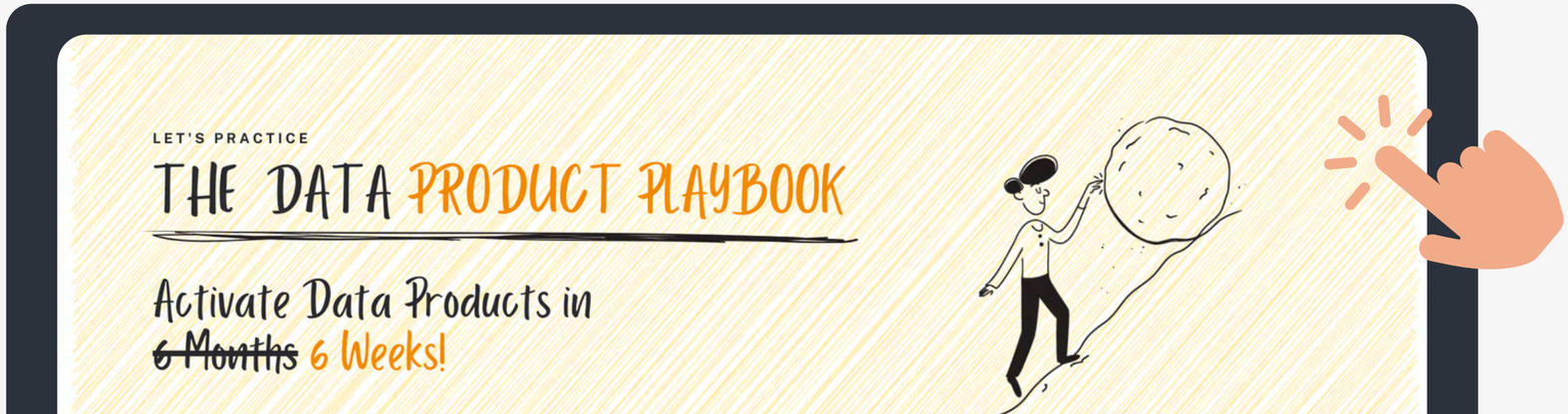
Enterprise AI is fundamentally a data product project with a model on top. Get the data product right, and the model follows.

The Data Product Playbook introduced the concept of treating data as a product, giving it ownership, a lifecycle, a defined consumer, and measurable value. That principle does not change for AI. In fact, it becomes more critical. AI models are data consumers. They are the most demanding, least forgiving, and highest-stakes data consumers your organisation has ever had to serve. A dashboard with slightly stale data is a minor inconvenience. A model trained on stale or inconsistent data makes decisions that can cost customers, revenue, and trust at scale.

The vision of this playbook is a three-layer architecture that connects your raw data to your AI outcomes.

This is a practical sequence. You start by building Data Products, which you already know how to do from the Data Product Playbook. In this playbook, you extend that practice specifically for AI: structuring your data products to serve as feature stores, training datasets, and real-time serving layers.

The self-serve platform that makes building, deploying, and evolving those data products fast enough to keep up with the pace of AI development.



The vision of this playbook is a three-layer architecture that connects your raw data to your AI outcomes.

Modern Data 101 Community

The Three-Layer Architecture

LAYER 1
PRODUCT FACTORY
(Self-Serve Infra)

Modular infrastructure components (workflows, compute, connectors, policies, monitors, contracts) assembled like LEGO blocks to power the data product layer at speed and scale

LAYER 2
DATA PRODUCTS

Feature stores, semantic models, domain datasets, and curated metric trees: versioned, documented, access-controlled, and tested. The AI-ready data layer.

LAYER 3
AI APPLICATIONS

Prediction services, recommendation engines, copilots, generative AI features, and decision intelligence: consuming reliable, governed data products.

THE PRODUCT FACTORY: SELF-SERVE DATA & CONTEXT FOR AI

Modern Data 101 Community

In Part I of this playbook series, we introduced the Self-Service Infrastructure (SSI) as the platform layer that enables data engineers to build data products declaratively, without managing every underlying resource from scratch. In the AI context, where scale is significant, this SSI becomes something more: consider something like a Product Factory.

This is essentially an SSI platform specifically capable of provisioning, orchestrating, and governing the full lifecycle of **AI-ready data products**: from raw data ingestion to feature computation, model serving pipelines, and feedback integration. All through modular, reusable infrastructure components.

> *The SSI is the infrastructure that makes your AI tools achieve significant measurable gains.*

For instance, the Data Operating System establishes a standard SSI for data productisation at scale, serving as a perfect underlying architecture that feeds AI applications.

90% reduction in time to deploy new AI feature

70% reduction in training / serving skew incidents

4X faster model iteration cycles

Read more on AI-Ready Data vs. Analytics-Ready Data

THE PRODUCT FACTORY: SELF-SERVE DATA & CONTEXT FOR AI

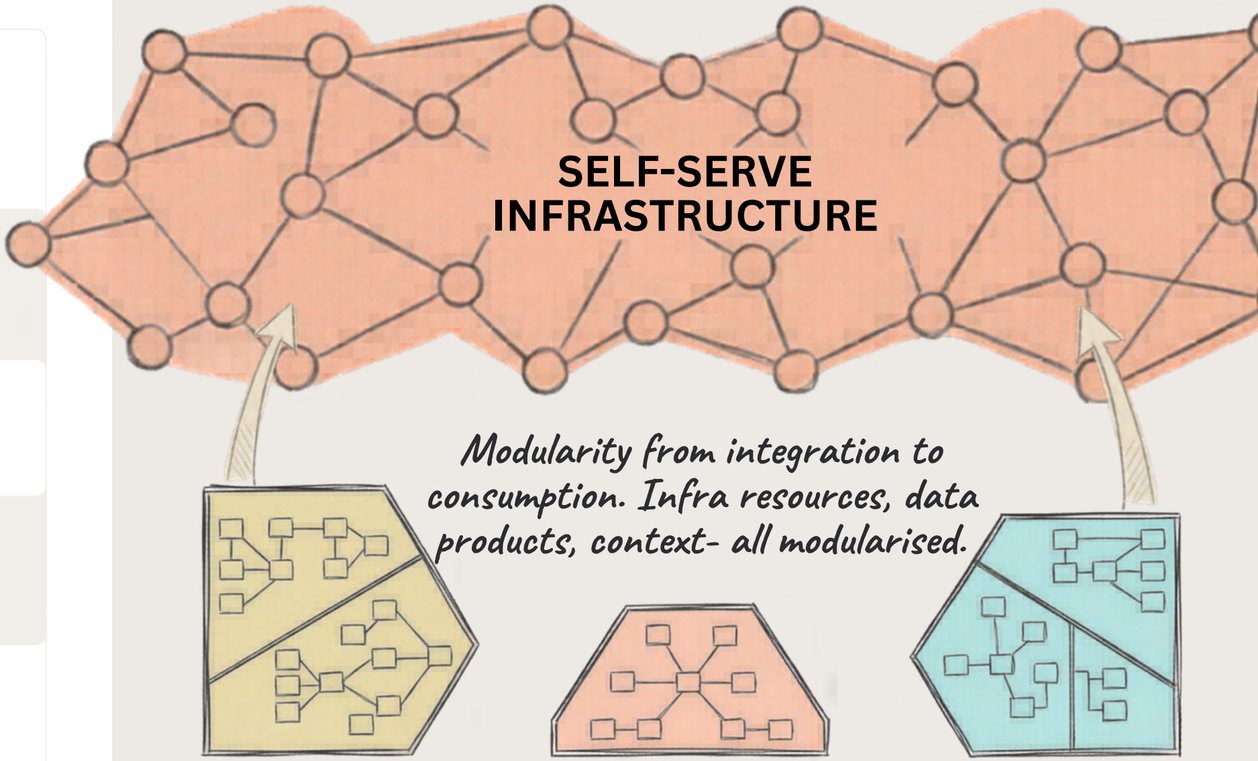
Modern Data 101 Community

WHAT MAKES IT A FACTORY?

In manufacturing, a factory does not produce one item. It produces many items of consistent quality, at speed, by combining standardised components in different configurations. A Product Factory for data does the same:

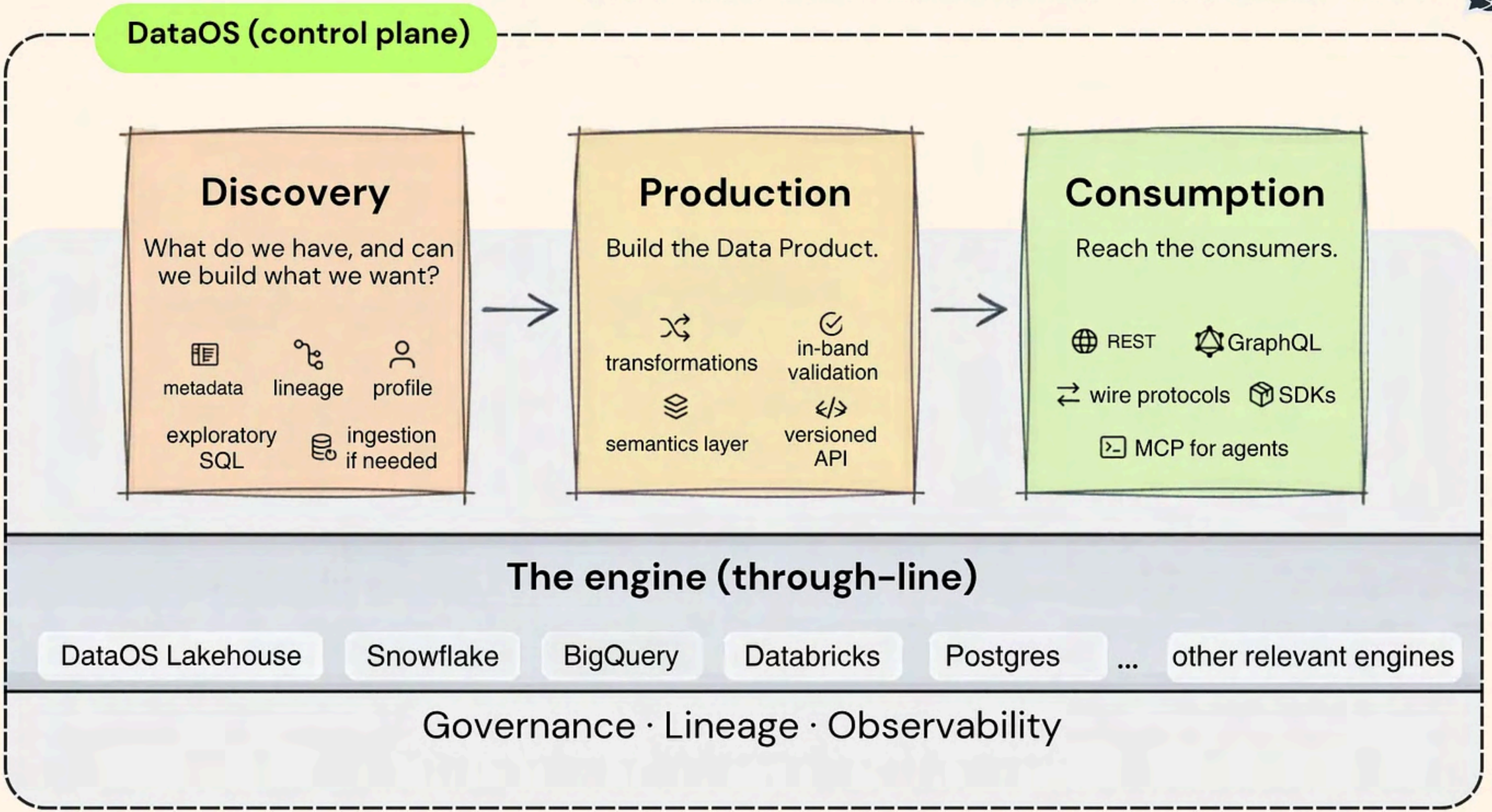
Modular Components	Reusable infrastructure primitives (like workflows, compute instances, data connectors, quality monitors, policy enforcers, serving endpoints) that can be combined in different configurations for different use cases.
Declarative Assembly	Engineers declare what they need (inputs, transformations, outputs, SLOs) and the factory provisions and wires together the underlying infrastructure automatically.
Standardised Quality	Every data product that comes out of the factory inherits the same baseline quality guarantees, governance controls, and monitoring, regardless of which team built it.
Speed at Scale	Because components are pre-built and reusable, building the 10th data product is nearly as fast as building the first. AI teams can experiment and iterate without waiting in infrastructure queues.
AI-Native Extensions	Beyond the core SSI, the Product Factory adds AI-specific components: feature computation pipelines, training dataset versioning, model-serving connectors, and inference monitoring.

Data Products
Indivisible Units of Context



BUILDING SELF-SERVE DATA FOR AI

ANIMESH KUMAR, SHUBHANSHU JAIN, DEVENDRA RATHORE X MODERN DATA 101



THE PRODUCT FACTORY: SELF-SERVE DATA & CONTEXT FOR AI

Modern Data 101 Community

The central approach of a standard self-serve infrastructure is LEGO-like blocks or architecture primitives. It deserves its own dedicated segment because it is the key to understanding why this approach is fundamentally different from how most laggard organisations are currently building their AI infrastructure.

Most enterprise AI infrastructure is built like a custom sculpture: every piece is bespoke, every integration is hand-crafted, every deployment is unique. The result is impressive until you try to build something slightly different and discover that **nothing reuses, nothing shares, and every new use case requires rebuilding from scratch.**

The LEGO block approach treats infrastructure components as standardised, connectable units. Any block can connect to any other block. New use cases are assembled from existing blocks and solutions.

The power of the LEGO block principle is combinatorial. For instance, nine core blocks combined in different configurations can serve dozens of use cases. Add nine AI-native extension blocks, and the number of possible configurations, and therefore possible AI data products, becomes extraordinary.

*You don't need more tools. You need **better building blocks that work together**. A Self-Serve Infrastructure provides the blocks. Your data product teams do the assembling.*

CORE INFRASTRUCTURE BLOCKS

- Workflow
- Service (API & microservice wrapper)
- Compute (scalable processing)
- Secret (credential management)
- Policy (access governance)
- Monitor (quality & drift detection)
- Contract (schema & SLA enforcement)
- Connector (source & sink adapters)

and more...

Read more on Self-serve infrastructure blocks

AI-NATIVE EXTENSION BLOCKS

- Feature Store (versioned feature sets)
- Training Dataset Builder (point-in-time correct joins)
- Embedding Pipeline (vector computation & storage)
- Model Serving Connector (ML platform integration)
- Inference Monitor (prediction drift detection)
- Feedback Ingester (outcome data back to features)
- Prompt Registry (LLM prompt versioning & governance)
- RAG Connector (retrieval-augmented generation setup)

WHERE ARE YOU IN THE ENTERPRISE AI JOURNEY?

Modern Data 101 Community

Understanding where you are is the prerequisite to knowing where to start.

Most enterprises reading this playbook are at Stage 2 or Stage 3. The 6-week framework in this playbook is designed to move you decisively toward Stage 4, with a clear path to Stage 5.

Stage 1

AI Experiments

You have proof-of-concepts running in notebooks. Data scientists are exploring models on extracted CSVs or ad-hoc queries. There is no production AI. The data layer is an afterthought.

Stage 3

Reactive AI Platform

You have invested in AI tooling: an ML platform, a vector database, maybe a feature store. But the data feeding these tools is still inconsistent, the feature logic is duplicated across teams, and the time to deploy new AI features is still measured in months.

Stage 5

Self-Serve at Scale

The full vision: a self-serve platform that enables any domain team to build AI-ready data products independently, with consistent quality and governance baked in. AI features are shipped as fast as business needs emerge. The data layer continuously improves from model feedback.

Stage 2

AI in Production

(Barely)

You have one or two AI models in production, maintained heroically by a small team. Every deployment is a crisis. Feature pipelines break regularly. Governance is minimal. Replicating this for more use cases feels impossible.

Stage 4

Data Product-Powered AI

Your AI features are backed by productised data: versioned feature sets with quality SLOs, governed access, and semantic documentation. New AI use cases are assembled from existing data products, with new ones built in weeks, not months.

BUILDING YOUR FIRST AI-READY DATA PRODUCT

Modern Data 101 Community

Start With a Real Business Problem

Do not start with a model. Start with a business outcome that AI could improve. 'Reduce churn by 15%' is a business problem. 'Train a churn prediction model' is not. The former drives the latter, never the reverse.

01

Treat Features as Data Products

Features are your most valuable AI asset. Every feature you build for one model is potentially reusable in five more. Productise them from day one: document them, version them, govern them, and register them in a feature catalog.

02

Build for Reuse, Not Just Deployment

Build for Reuse, Not Just Deployment
Every data product you build for AI should be designed with the question: who else in the organisation could use this? The marginal cost of serving a second consumer from a well-built data product is near zero.

03

Close the Loop from Model to Data

Every AI model generates signals about data quality that most organisations ignore. When a model's predictions degrade, it is almost always because the data has changed. Build feedback mechanisms from your models back to your data products from the start.

04

BEFORE YOU BEGIN: APPOINT AN AI PRODUCT LEAD

Modern Data 101 Community

In Part I, we recommended assigning an Acting Data Product Manager to lead your 6-week proof of value. For the Enterprise AI Playbook, you need that person and one more: an AI Product Lead. The **AI Product Lead** sits at the intersection of AI/ML engineering, data product management, and business domain expertise. This is not a data scientist who occasionally talks to business stakeholders. This is someone who can translate between model performance metrics and business KPIs, who understands feature engineering as a product discipline, and who can prioritise AI use cases by business impact rather than technical novelty.

DATA PRODUCT MANAGER

- Owns the semantic model and metric tree
- Manages data consumer relationships
- Governs data product quality SLOs
- Drives adoption within business domains
- Prioritises data product roadmap

AI PRODUCT LEAD (NEW)

- Translates business problems to AI use cases
- Owns feature store strategy and catalog
- Connects model performance to data quality
- Manages training/serving consistency
- Closes the feedback loop from model to data

You may not need to hire for this role immediately. Look for a senior ML engineer or data product manager who already thinks this way. Assign them as the AI Product Lead for your 6-week proof of value. If it succeeds, and it will, make the role permanent.

WEEK 1

AI USE CASE DISCOVERY & DATA FITNESS AUDIT

FIND THE AI OPPORTUNITY. ASSESS YOUR DATA'S READINESS TO SERVE IT.

Modern Data 101 Community

The most common mistake in enterprise AI is starting with the model. The second most common is starting with the data. The right place to start is with the business problem, and then working backwards to understand what data is required to solve it, and whether that data is fit for purpose.

Week 1 is a discovery week. You will conduct structured interviews with business stakeholders, map the AI use cases that have the highest potential business impact, and run a data fitness audit that tells you honestly whether your current data estate can support those use cases.

AI Use Case Discovery

Run structured sessions with domain stakeholders, like the Sales Director, the Head of Marketing, the VP of Customer Success, and ask them the following:

- Where do your team's decisions currently rely on where data could help?
- What is the business outcome you are most trying to move in the next 12 months?
- Where do you see delays, errors, or missed opportunities that repeat in a predictable pattern?
- What would you do differently if you had a reliable signal 24 hours earlier?
- What does 'good prediction' or 'good recommendation' look like in your domain?

WEEK 1

AI USE CASE DISCOVERY & DATA FITNESS AUDIT

FIND THE AI OPPORTUNITY. ASSESS YOUR DATA'S READINESS TO SERVE IT.

Modern Data 101 Community

The AI Use Case Scoring Matrix

Not all use cases are worth pursuing first. Score each identified use case across four dimensions and prioritise the highest-scoring opportunities for your 6-week proof of value:

DIMENSION	WHAT TO ASSESS	SCORE (1-5)
Business Impact	Direct connection to revenue, retention, or cost KPIs	High impact = 5
Data Availability	DIs the required data already collected and accessible?	Ready now = 5
Data Quality	Is the data clean enough for a model to learn from?	High quality = 5
Stakeholder Commitment	Is the domain leadership willing to act on predictions?	Strong buy-in = 5

WEEK 1

AI USE CASE DISCOVERY & DATA FITNESS AUDIT

FIND THE AI OPPORTUNITY. ASSESS YOUR DATA'S READINESS TO SERVE IT.

Modern Data 101 Community

The Data Fitness Audit will almost always reveal gaps. That is its purpose. The gaps you find in Week 1 define the data product work you need to do in Weeks 2 and 3.

The Data Fitness Audit

For your top-scoring use case, run a structured audit of the data required to power it. This is not a technical deep-dive; it is a product-readiness assessment:

Availability

Does the data exist? Is it being collected today? Where does it exist?

Freshness

How current is it? What is the latency from event to availability? Is that sufficient for the use case?

Completeness

What percentage of records have the fields needed? What is the historical depth?

Consistency

Is the same entity defined the same way across all sources? Are there naming conflicts or semantic ambiguities?

Lineage

Do you know where this data came from and how it was transformed?

Access

Can the AI team access this data without a multi-week access request process?

Governance

Are there compliance or privacy constraints that affect what can be used for model training?

WEEK 2

FEATURE ENGINEERING AS A DATA PRODUCT

THE CANVAS: DESIGNING YOUR AI-READY SEMANTIC MODEL AND FEATURE SETS.

Modern Data 101 Community

Feature engineering is where AI projects are won or lost. A good feature is the distilled intelligence of a domain expert, encoded in a form that a model can learn from. Week 2 extends the Data Product Canvas specifically for AI use cases. You will build on your semantic model and extend it with feature-level entities that your models need: governed, versioned, and catalogued from day one.

Extending the Semantic Model for AI

Your semantic model already defines business entities, dimensions, and measures. For AI, you add a fourth layer: Features, derived, transformed, or aggregated representations of your entities that are specifically designed to be predictive.

Entities

Core business objects: customer, transaction, product, session, deal (from your existing semantic model)

Dimensions

Aggregated quantitative facts: revenue, session_count, avg_order_value (from your existing model)

Measures

Descriptive attributes of entities: tenure, region, plan_type, industry (from your existing model)

Features (NEW)

Derived predictors: 30_day_engagement_score, days_since_last_purchase, rolling_churn_signal, support_ticket_velocity. Engineered specifically for model training and inference.

Labels (NEW)

Historical outcomes that the model is trained to predict: churned_30d, converted, high_value_segment, at_risk_flag. Requires careful point-in-time construction to avoid leakage.

Feature Groups (NEW)

Logical groupings of related features that are versioned, documented, and served together: e.g., 'Customer Engagement Features v2.1'.

FEATURE ENGINEERING AS A DATA PRODUCT

THE CANVAS: DESIGNING YOUR AI-READY SEMANTIC MODEL AND FEATURE SETS.

Modern Data 101 Community

Point-in-Time Correctness: The Non-Negotiable

Your semantic model already defines business entities, dimensions, and measures. For AI, you add a fourth layer: Features, derived, transformed, or aggregated representations of your entities that are specifically designed to be predictive.

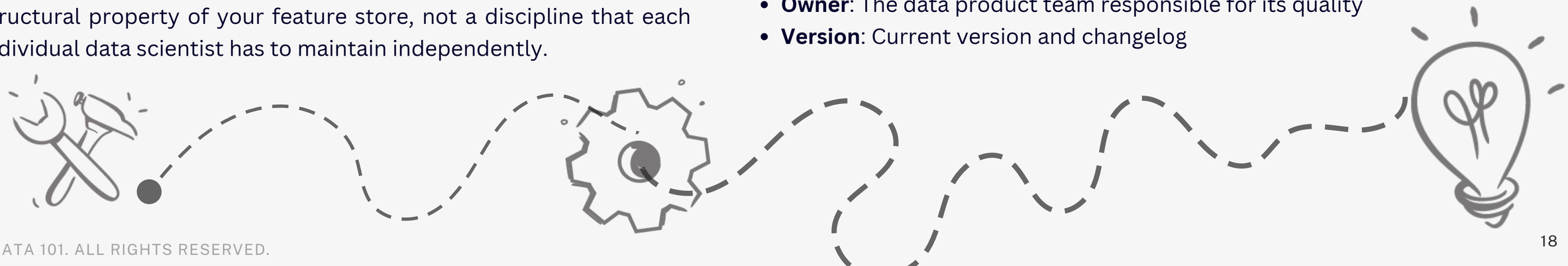
> *The most common source of model failure in production is training-serving skew. The leading cause of training-serving skew is using feature values as they exist today to train a model on historical labels. Always reconstruct features as they existed at the time of the label event.*

When you build features as data products with explicit versioning and timestamp semantics, point-in-time correctness becomes a structural property of your feature store, not a discipline that each individual data scientist has to maintain independently.

The Feature Catalog: Your AI Asset Registry

Every feature you engineer should be registered in a Feature Catalog, the AI equivalent of a data catalog. Each entry in the catalog includes:

- **Feature Name**: Human-readable, domain-prefixed (e.g., customer.engagement_score_30d)
- **Definition**: Plain-language description of what this feature represents
- **Computation Logic**: The formula or transformation that produces it
- **Update Frequency**: How often it is recomputed (hourly, daily, on-event)
- **Freshness SLO**: The maximum acceptable staleness for a consumer
- **Consuming Models**: Which models currently use this feature
- **Owner**: The data product team responsible for its quality
- **Version**: Current version and changelog



WEEK 3

AI PIPELINE ACTIVATION

THE BUILD: CONNECTING YOUR DATA PRODUCT LAYER TO MODEL TRAINING, SERVING, AND MONITORING.

Modern Data 101 Community

This is where the Data Product Factory earns the returns. Without a self-serve platform, Week 3 would take months: custom pipeline code, bespoke integrations, manual credential management, and one-off deployment scripts that only the author can maintain. With the self-serve infrastructure (SSI), your analytics engineer writes a single declarative specification and the platform provisions what is needed.

THE FOUR AI PIPELINE COMPONENTS

TRAINING PIPELINE

Orchestrates the creation of a training dataset from your feature store. Handles point-in-time joins, label generation, train/validation/test splits, and dataset versioning.

Output: a versioned, reproducible training dataset registered in the feature catalog.
SSI resources used: Workflow, Feature Store, Dataset Builder, Secret

SERVING PIPELINE

Computes features in real-time or near-real-time for inference. Ensures that the feature logic used at serving time is identical to the logic used at training time, eliminating the training-serving skew problem structurally.

Output: low-latency feature vectors served to your model inference endpoint.
SSI resources used: Service, Compute, Feature Store (online), Monito

QUALITY & GOVERNANCE PIPELINE

Continuously validates the data flowing through your AI pipeline against your SLOs, checking for schema drift, distribution shift, missing values, and freshness violations. Triggers alerts and, where configured, automatic fallback behaviour.

Output: a dynamic quality dashboard and automated SLO enforcement.
SSI resources used: Monitor, Contract, Policy, Workflow

FEEDBACK PIPELINE (NEW)

Captures prediction outcomes and actual results and routes them back into the data product layer as new labels or feature signals. Enables the data product to improve continuously from model performance.

Output: a dynamic dataset of predictions vs. outcomes, feeding the next training cycle.
SSI resources used: Feedback Ingester, Workflow, Feature Store, Contract

AI PIPELINE ACTIVATION

Modern Data 101 Community

THE BUILD: CONNECTING YOUR DATA PRODUCT LAYER TO MODEL TRAINING, SERVING, AND MONITORING.

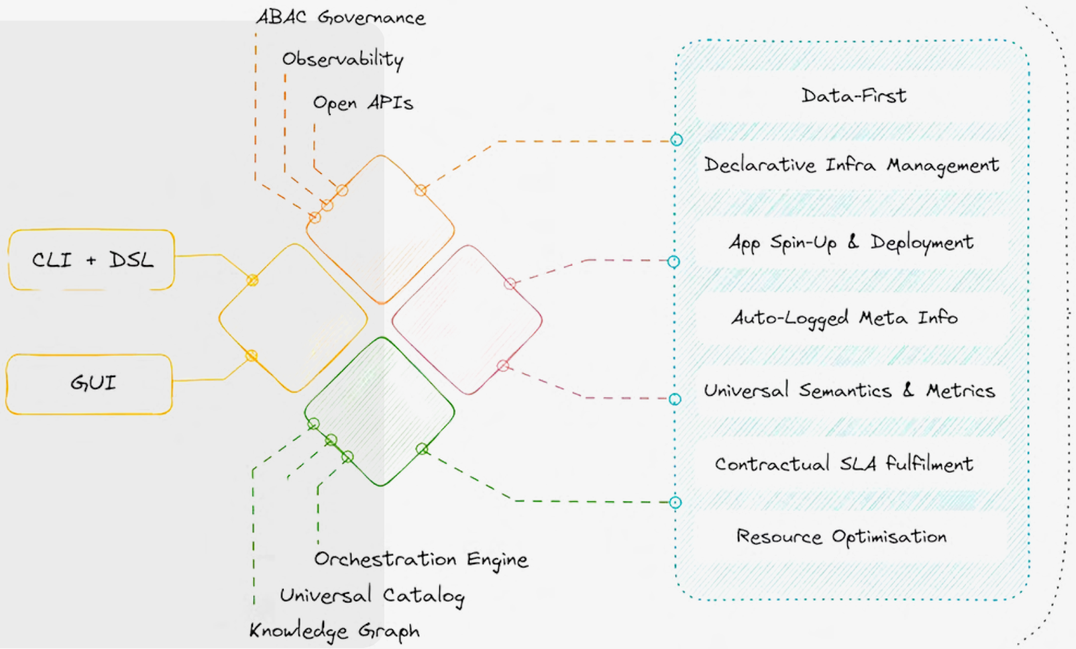
Declarative Activation: The Developer Experience

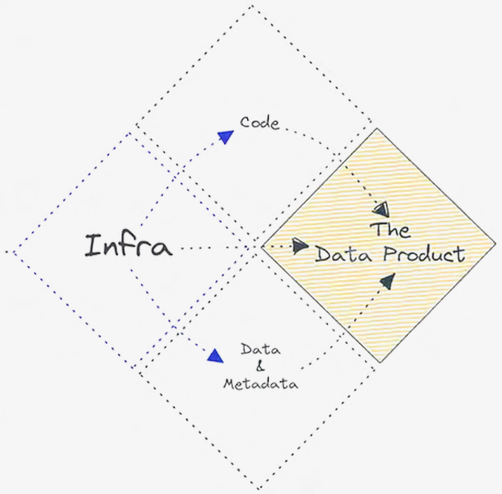
Your analytics engineer activates all four pipeline components through a declarative specification, describing inputs, transformation logic, outputs, and quality conditions without hand-coding the infrastructure:

The self-serve infrastructure (SSI) reads this specification and provisions the training dataset pipeline, including feature computation, point-in-time join logic, quality monitoring, and dataset versioning automatically.

Sample declarative spec (YAML): AI Training Pipeline

```
name: churn-prediction-training-dp
type: ai-training-dataset
feature_groups:
 - customer.engagement_features_v2
 - customer.support_signal_features_v1
label: customer.churned_30d
point_in_time: true
lookback_days: 365
slos:
 completeness: '>= 95%'
 freshness: '<= 24h'
```





WEEK 4

AI PRODUCT GO-TO-MARKET

THE LAUNCH: ADOPTING YOUR AI DATA PRODUCT ACROSS THE DOMAIN.

Modern Data 101 Community

The most technically impressive AI product is worthless if business users don't adopt it. Week 4 is about ensuring your AI-ready data product and the model that consumes it actually change how the domain team makes decisions.

AI adoption fails most often not because the predictions are wrong, but because the business team was never meaningfully involved in defining what 'right' means, or because the predictions aren't surfaced in the workflow where decisions are actually made.

THE AI PIPELINE COMPONENTS

PILLAR 1
Explanation

Business users trust predictions they understand. Surface not just the prediction score but the top features driving it. If a customer is predicted to churn, the Sales Manager needs to know why: 'declining login frequency + 2 unresolved tickets + contract renewal in 14 days' is actionable. '0.87 churn probability' not so much.

PILLAR 2
Embed in Existing Workflows

Do not build a new dashboard that users have to navigate to separately. Embed predictions where decisions are made: in the CRM, in the ticketing system, in the daily standup report. The best AI product is invisible; it just makes existing workflows smarter.

PILLAR 3
Measure Adoption as Rigorously as Accuracy

Track not just model performance metrics but adoption metrics: What % of at-risk customers identified by the model received an intervention? What was the conversion rate of model-guided outreach vs. unguided outreach? Adoption metrics are your ROI proof.

WEEK 4

AI PRODUCT GO-TO-MARKET

THE LAUNCH: ADOPTING YOUR AI DATA PRODUCT ACROSS THE DOMAIN.

Modern Data 101 Community

AI Product Positioning

Frame the prediction in the language of the domain: not 'churn probability score' but 'Retention Risk Alert'. Map every output of the model to a specific decision or action the domain team can take.

Champion Identification

Find 2-3 power users in the domain who will test the product first and advocate for it internally. Their word-of-mouth is more valuable than any internal communication.

Feedback Mechanisms

Give domain users a way to mark predictions as correct, incorrect, or 'can't determine'. This feeds directly back to your model retraining and data quality improvements.

Success Metrics

Define what 'success' looks like for the 6-week proof of value in business terms (churn reduced, revenue saved, interventions that converted).

WEEK 5 & 6

MODEL PERFORMANCE & CONTINUOUS FEEDBACK

THE PROOF: MEASURING VALUE, CLOSING THE LOOP, AND PREPARING TO SCALE.

Modern Data 101 Community

The final two weeks of your 6-week proof of value are the most important and the most commonly rushed. This is where you close the loop between AI output and data input, validate the business impact of your proof of value, and build the foundation for scaling from one AI use case to many.

A TALE OF TWO DASHBOARDS

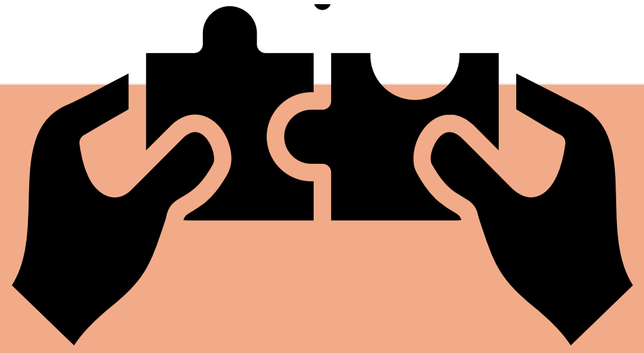
By Week 5, you should be running two dashboards simultaneously:

MODEL PERFORMANCE DASHBOARD

- Precision, recall, and F1 by segment
- Prediction confidence distribution
- Feature importance rankings
- Data quality SLO compliance rate
- Training-serving feature drift alerts
- Label distribution over time

BUSINESS IMPACT DASHBOARD

- Interventions triggered by model predictions
- Conversion rate: model-guided vs. baseline
- Revenue impact: retained accounts, saved deals
- Adoption rate: % of predictions acted upon
- User feedback scores on prediction quality
- Business metric movement (North Star KPI)



WEEK 5 & 6

MODEL PERFORMANCE & CONTINUOUS FEEDBACK

THE PROOF: MEASURING VALUE, CLOSING THE LOOP, AND PREPARING TO SCALE.

Modern Data 101 Community

THE FEEDBACK LOOP ARCHITECTURE

The feedback loop is the mechanism that makes your AI data product self-improving over time. Without it, your model is static. It learns from historical data and then stops. With it, every real-world prediction and its outcome becomes a new training signal.

01

Capture Outcomes

Record what actually happened after each prediction. Did the at-risk customer churn? Did the flagged fraud transaction turn out to be fraud?

02

Enrich Feature Store

Route outcome data back into the feature store as new label data, enabling continuous model retraining on fresh, real-world examples.

03

Detect Data Drift

Use your Monitor blocks to detect when the distribution of incoming feature values diverges from the training distribution: a leading indicator that model performance is about to degrade.

04

Trigger Retraining

Configure your Product Factory to trigger automatic retraining when drift exceeds a defined threshold, or at a regular cadence based on the dynamics of your domain.

05

Evolve SLOs

As you learn more about how the model and data interact, refine your data quality SLOs: tighten freshness requirements, add completeness checks, enforce schema contracts more strictly.

WEEK 5 & 6

MODEL PERFORMANCE & CONTINUOUS FEEDBACK

THE PROOF: MEASURING VALUE, CLOSING THE LOOP, AND PREPARING TO SCALE.

Modern Data 101 Community

THE FEEDBACK LOOP ARCHITECTURE

The 6-week proof of value proves the concept. The architecture you built proves the platform. Now the question is: how quickly can you replicate this for the next AI use case?

The answer depends almost entirely on how well you productised your data layer in Weeks 2 and 3. If your features are documented, versioned, and catalogued, a new use case that shares feature groups with the first is a matter of configuration. This is the compounding return on the Product Factory investment.

Use Case Expansion

Map which of your remaining AI use cases can reuse features you have already built. Prioritise these first: the ROI on the second use case is dramatically higher than the first.

Feature Reuse Rate

Track the percentage of features in new AI products that are reused from existing feature groups. This is your Product Factory efficiency metric. A healthy target is >60% after your third use case.

Platform Contribution

Document each new infrastructure component added during your proof of value and contribute it to the shared Product Factory as a new reusable block.

GOVERNANCE, TRUST & RESPONSIBLE AI

Modern Data 101 Community

AI governance is not a compliance checkbox. It is the foundation of trust, and trust is the prerequisite for adoption. An AI product that domain users do not trust will not be used, regardless of its predictive accuracy.

The good news is that if you have built your AI products on a productised data layer with the governance controls described in this playbook, you are already most of the way there. The governance capabilities that make data products trustworthy are the same ones that make AI products trustworthy.

THE FIVE GOVERNANCE PILLARS FOR ENTERPRISE AI

Data Lineage

Every prediction made by your model should be traceable back through the feature computation to the raw data sources.

When a prediction is wrong, you need to know whether it is a model problem or a data problem, and lineage tells you. The Self-serve Infrastructure (SSI) Contract and Monitor blocks provide this automatically.

Feature Governance

Features used in production models must be version-controlled, with changes requiring review and approval.

A silent change to a feature definition is equivalent to retraining the model on different data without knowing it

Access Controls

Not all business users should have access to all predictions, particularly in regulated industries.

The SSI's Policy blocks enforce row-level, column-level, and use-case-level access controls on your AI data products.

Fairness Monitoring

Models trained on historical data can perpetuate or amplify historical biases.

Configure your inference monitoring to track model performance broken down by demographic segments, and set alerts when performance diverges significantly across groups.

Explainability Standards

For high-stakes decisions (credit, healthcare, hiring, enforcement) set explicit standards for prediction explainability.

The model should be able to produce a plain-language explanation of every prediction it makes.

GOVERNANCE, TRUST & RESPONSIBLE AI

Best Practices

Modern Data 101 Community

Embed these practices into your data product development process from the get go:

- Conduct a data ethics review before productising any dataset for AI training. Check for potential sources of bias, historical discrimination, and protected attributes.
- Define the 'human in the loop' protocol for every AI product: which decisions require human review before action, and how does the prediction UI support that review?
- Establish model cards for every production model: a standardised document describing the model's intended use, performance characteristics, known limitations, and governance contacts.
- Set a model deprecation policy. When is a model retired? What is the process for communicating deprecation to downstream consumers?
- Maintain an AI incident register. When a model makes a decision that causes harm or significant error, document it, investigate it, and improve the data product accordingly.

The organisations that will scale enterprise AI fastest are not those with the most powerful models. They are those with the most trustworthy data products underneath them.

Modern Data 101 Community

C A N V A S

Enterprise AI Readiness Canvas

Map your AI initiative from business problem to production feedback loop: left to right.

Use Case / Domain:	AI Product Lead:	Date:	Stage (1-5):

JOURNEY LEFT-TO-RIGHT Start with the business problem. Work right through data readiness, the product factory, the AI-ready data layer, and into the application. Then close the loop.

Business Domain	Semantic Model (Extended)	AI Data Products (Feature Store)	Model & Serving	AI Application Layer
• What business outcome (revenue, retention, cost) are we moving? • Who owns the decision that this AI use case informs? • Stakeholder commitment secured?	• Entities, dimensions, measures already defined? New Features & Labels needed for this use case? • Point-in-time correctness designed in?	• Feature groups versioned & catalogued? • Freshness / completeness / consistency SLOs set? • Reusable by a second use case?	• Training/serving pipelines use identical feature logic? • Inference monitoring & drift detection exist? • Explainability output attached to every prediction?	• Embedded in the workflow where decisions are made? • Champion users identified for adoption? • Intervention playbook documented per prediction?
→ Start here	Feedback Loop→→→→→→→→→→→→			

Product Factory (Self-Serve Infrastructure): The LEGO blocks underneath every column above
Workflow. Service. Compute. Connector. Secret. Feature Store. Monitor. Contract. Policy. Feedback Ingest. and so on...

Data Fitness Audit	Governance & Trust	People & Roles	Success Metrics
• Availability: Is the data available, and where? • Freshness: Latency from event to availability • Completeness: % records with required fields • Consistency: Same entity, same definition • everywhere • Lineage: Do we know where it came from? • Access: Can the AI team reach it without a • queue?	• Data lineage is traceable for every prediction • Feature changes version-controlled & reviewed • Access controls (row / column / use-case level) • Fairness monitoring across segments • Explainability standard for high-stakes • decisions • Model card & deprecation policy on file	• AI Product Lead appointed • Data Product Manager aligned • Domain champion(s) identified (2-3 users) • Platform / ML engineering support confirmed • Executive sponsor for the 6-week proof of value	• Business KPI this use case must move • Model metric (precision / recall / F1) target • Adoption rate % predictions acted upon • Feature reuse rate for the next use case • Time-to-deploy for use case #2 vs. #1

Where are you starting from? Mark your current stage - most enterprises begin at Stage 2 or 3. This canvas is designed to move you to Stage 4.

STAGE 1 AI Experiments	STAGE 2 AI in Production (Barely)	STAGE 3 Reactive AI Platform	STAGE 4 Data Product-Powered AI	STAGE 5 Product Factory at Scale

Join the Modern Data 101 Crusade!

If this playbook was helpful for your implementation or even as a means to organise your thoughts, we'd highly recommend you join us and help us build data product and AI-readiness expertise in the data industry.

Find a truly community-driven voice with experts from all walks of the data industry voicing their insights on Data & AI across Modern Data 101 channels. We'd love for you to come forward and share your insights with us so we can improve our delivery and truly create for YOU.

Subscribe

20,000+
STRONG
A growing community of data and AI leaders

130+
COUNTRIES
From every corner of the world

150+
CONTRIBUTORS
Sharing knowledge, ideas and experience

400+
RESOURCES
Curated learning on data and AI

Be part of something bigger.

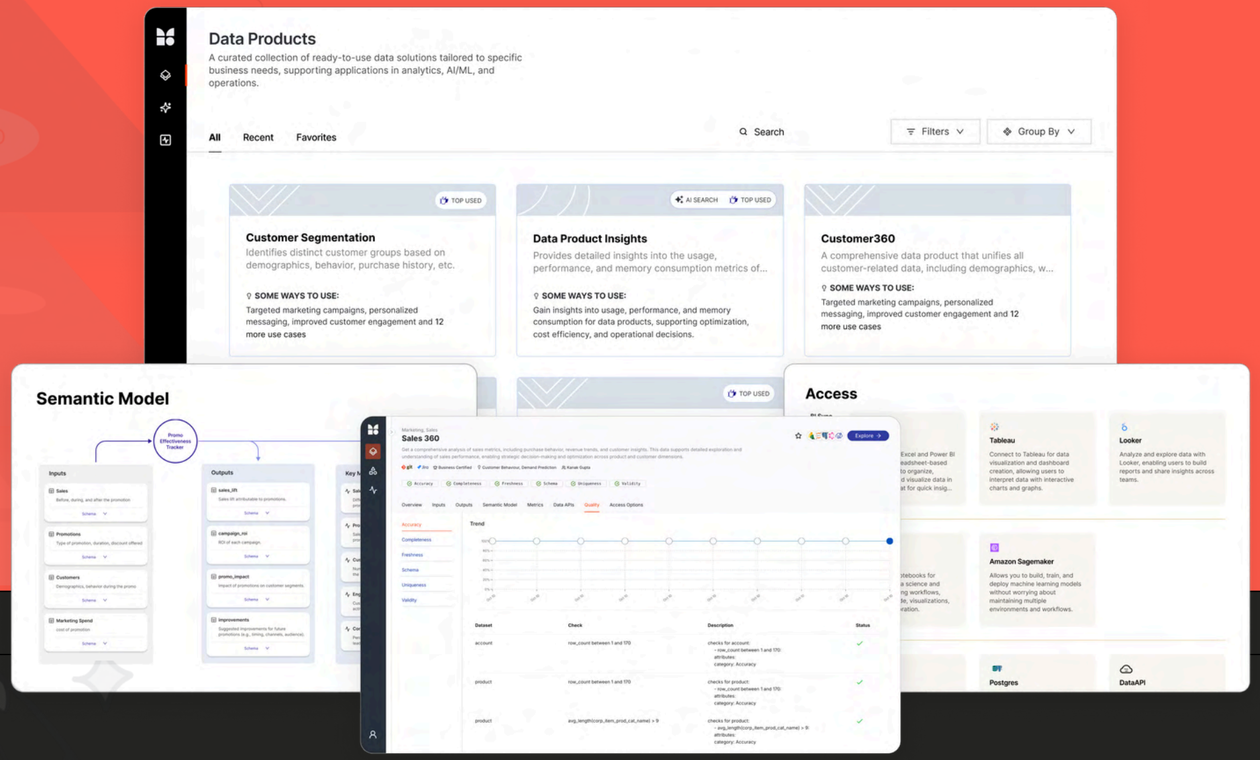
Whether you're just starting out or a seasoned pro, there's a place for you here.

Let's build the future of data and AI—together.

Study Facilitated by The Modern Data Company

The Modern Data Company was founded to radically simplify how organizations manage, access, and interact with data. Modern's data operating system, DataOS®, unifies all data across silos, at scale. The composable layers of DataOS evolve with your existing technology stack to future-proof your data initiatives.

themoderndatacompany.com info@tmdc.io



Study Conducted by Modern Data 101

Modern Data 101 is a publication and community for people building data platforms, designing data teams, and architecting the invisible. In a world of countless tools, trends, and templated thought leadership, Modern Data 101 slows things down. We ask: Why was this built? Each piece is a lens. On architecture. On semantics. On organisational design.

moderndata101.com community@moderndata101.com