Unframe

The New Rules for Managing Data with AI

unframe.ai

Unframe

Table of contents

Unframe

Key takeaways

The AI data shift is massive and accelerating

The global AI market has surpassed $757 billion in 2025, up from $638 billion in 2024, and is projected to reach $2.74 trillion by 2032. This growth demands AI-first data systems built for intelligence, not just analytics.

From dashboards to AI readiness

Legacy architectures optimized for reporting must evolve into AI-ready ecosystems that support continuous learning, reasoning, and autonomous operations.

Automation and feedback loops as core design

Continuous automation and AI feedback loops power self-improving data systems that can detect drift, optimize cost, and enforce governance without manual intervention.

The new operating charter for data teams

Data engineers must become platform enablers, delivering governed, reusable, discoverable data products that drive enterprise-wide AI adoption.

The four-step framework for AI-first data management

1. Make data quality autonomous
2. Embed governance by design
3. Optimize for cost and scale automatically
4. Shift from management to enablement

Executive summary

AI is redefining enterprise data, and it's exposing the biggest barrier to AI readiness: the data itself. Systems built for static reporting can't deliver the dynamic, trustworthy, machine-readable data that AI needs to learn and act on its own.

The global AI market surpassed $638 billion in 2024. This year, it's towering past $757 billion, and by 2032 it's expected to climb all the way to $2.74 trillion. These figures are based on global and US data by Resourcera. With such growth, data engineering leaders face an inflection point. AI systems no longer consume data - they depend on it to learn, reason, and act autonomously. Traditional data management models can't keep up.

Still, too many organizations continue to run legacy data engineering operations built for dashboarding and analytics, not AI. Manual staging, quarterly audits, and batch governance break under the demands of hundreds of real-time models and agents.

You can use this 4-step framework, designed for enterprise data engineering leaders, to shift from "managing data" to "enabling data for AI outcomes". With each step, we provide practical explanation, flows, key capabilities, and real-world examples from companies already putting it into action.

The mindset shift from management to enablement

Before we dive into the steps, it's useful to frame the shift in mindset and operating model.

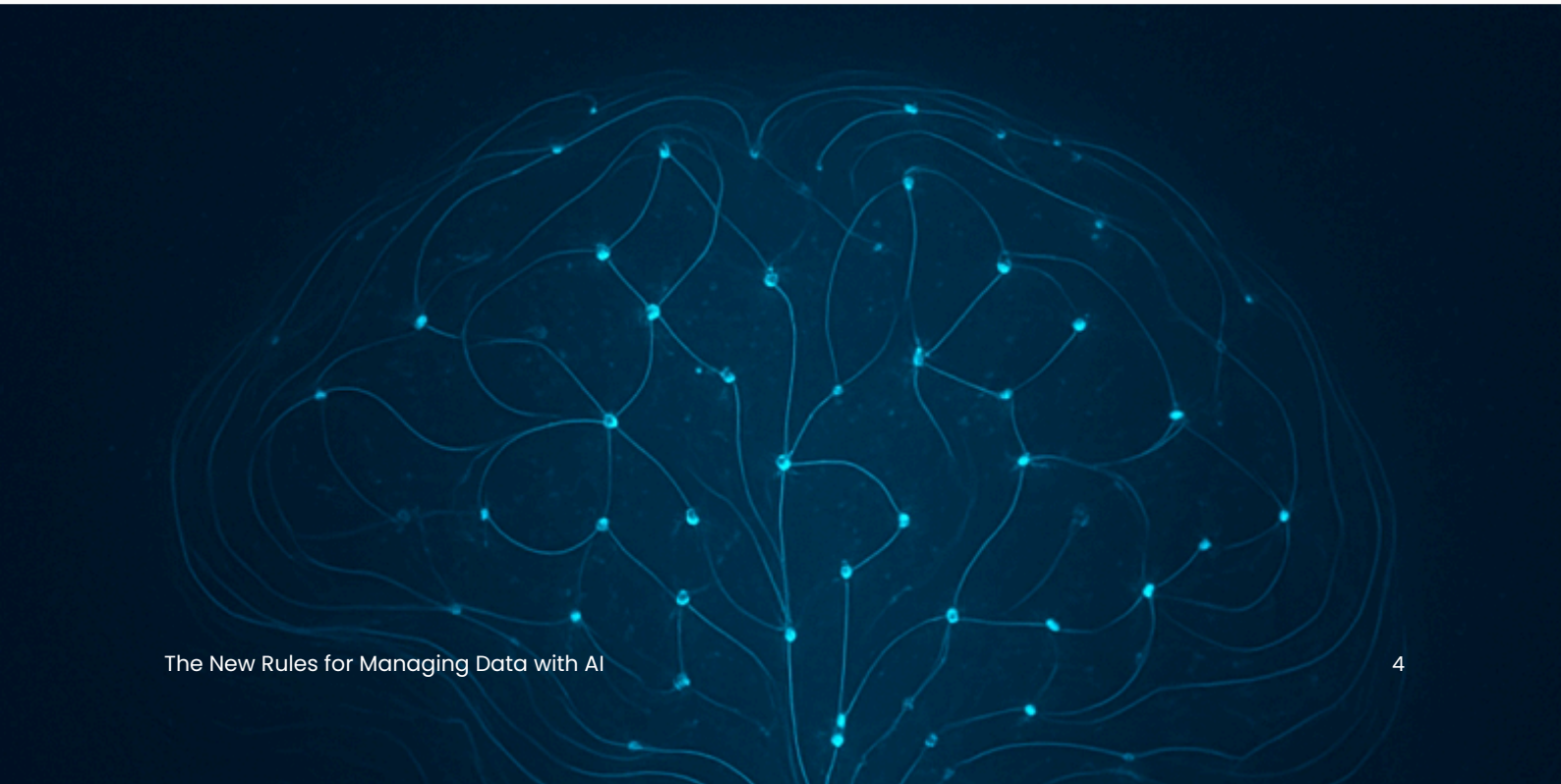
Old data engineering paradigm

- Focused on ETL/ELT, warehouses, and dashboards.
- Relied on manual quality checks and ad-hoc governance.
- Measured success by data volume and refresh latency.
- Scaled by adding headcount or hardware.

New AI-era paradigm

- Data must feed models, agents, and digital twins autonomously.
- Quality is continuous, governance is embedded, scale is self-optimizing.
- Success metrics shift to AI readiness, data reuse, and data-to-model speed.
- Scale comes from automation, platformization, and self-service.

> Global forecasts estimate **AI will add $15.7 trillion** to the world economy by 2030. The differentiator? Which enterprises can **operationalize their data for AI now**.



Unframe

The new operating model for AI-ready data engineering

To succeed, data engineering leadership needs to adopt four new rules. Below, we map out each rule as a step, describe flows and capabilities, and demonstrate with **real-world wins.**

The 4 Steps of the AI-First Data Framework

1

Make Data Quality Autonomous

Create self-healing pipelines that detect schema drift, bias, and data quality issues automatically. Replace manual checks with AI-driven, feedback-based systems that maintain real-time accuracy and trust.

2

Embed Governance by Design

Implement policy-as-code, lineage automation, and audit-ready logging so governance is built into every data pipeline, transforming compliance from a bottleneck into a continuous process.

3

Optimize for Cost & Scale Automatically

Deploy predictive autoscaling and AI-driven telemetry to monitor workload patterns, reduce waste, and allocate compute intelligently—achieving up to 30 % cost savings and faster performance.

4

Shift from Management to Enablement

Empower teams with governed self-service data catalogs and AI copilots that guide usage and discovery. This model replaces gatekeeping with enablement, doubling innovation speed and accelerating AI adoption.

Unframe

Step 1: Make data quality autonomous

AI systems don't tolerate stale, mis-labelled or inconsistent data. In fact, many failures of AI initiatives trace back to data quality issues like missing context, semantic mismatches, pipeline breaks, and biases creeping in. Build feedback-driven, self-healing pipelines that detect schema evolution, drift, and bias without waiting for human intervention.

What a flow looks like

1. Ingestion and monitoring:
Raw data flows from sources into the ingestion layer (streaming/batch). Metadata is captured at ingestion (schema, provenance, timestamp, domain).

2. Continuous quality engine:
An AI engine monitors quality metrics (completeness, consistency, drift, anomaly, freshness). When thresholds are breached, it raises flags or initiates remediation.

3. Feedback loops:
Downstream model outcomes, user corrections, data-operations metrics feed back into the quality engine. The system learns what "good" means for each domain.

4. Self-healing pipelines:
Upon detecting root causes (source schema changed, missing domain mapping), automated workflows kick off fix tasks — rerun jobs, apply imputations, alert domain owners.

5. Governance integration:
Quality metrics are fed into dashboards and audit logs; data engineering and data science teams review exceptions as part of governance.

Key capabilities

- AI anomaly detection of incoming data streams
- Metadata-driven monitoring with schema change detection, lineage triggers
- Feedback integration with model performance back into data quality engine
- Automated remediation workflows with alerting + tickets
- Quality SLA dashboards for data consumers and engineering ops
- AI-ready metadata to improve data discovery and accessibility for users

Real world success

S&P Global

S&P Global launched its "AI-Ready Metadata" initiative, making financial datasets machine-readable and enriched with context so that AI systems (in addition to humans) become primary consumers of the data which can then be queried in natural language by financial analysts.

Rather than handing data to analytics teams, S&P Global conditions data for AI consumption by embedding semantics, units, and synonyms, making metadata machine-actionable and self-describing. This is the hallmark of autonomous data quality at scale.

They achieved this by embedding meaning at the column level (units, relationships, cross-references, semantic tags) and exposing it through vendor-neutral APIs and Snowflake® distribution. This engineering decision eliminates one of the biggest friction points in the machine-learning lifecycle: the time-consuming pre-processing and normalization that delay model readiness.

Source: S&P Global

Walmart

Walmart's data engineering team rebuilt its retail data foundation to power the company's Element AI platform, which supports more than 900,000 associates and handles over 3 million daily queries. The team used generative AI to standardize 850 million product data points across its global catalog, reducing inconsistencies, removing duplicate identifiers, and improving data quality for all downstream AI workloads.

The engineering effort included automating ingestion pipelines across thousands of store and sensor systems, as well as building an AI-driven anomaly-detection layer that continuously monitored data freshness and correctness. Each correction - whether triggered by store associates, fulfillment telemetry, or customer-search behavior - feeds back into the pipeline, training the quality models that drive remediation.

For data engineering leaders, the lesson is clear: data quality must evolve from a periodic audit into a real-time, agentic feedback system.

Source: CX Dive

What this means

- Shift KPIs from "number of bad records corrected" to "percentage of datasets passing automated quality checks within minutes of ingestion".
- Establish dedicated 'data-quality engine' capability inside the data platform (not just a governance committee).
- Build instrumentation at ingestion that captures lineage, metadata, schema evolution and propagates into the automation layer.
- Plan for self-healing pipelines: when an upstream schema change occurs, how fast can your system detect, rerun and validate downstream flows?

Step 2: Embed governance by design

In AI-first enterprises, governance must be embedded in every pipeline, platform and data workflows. Governance by design means policy becomes code, lineage is traceable, and trust is built into every data product.

What a flow looks like

1. Policy definition as code:
Data engineers define policies (data retention, sensitivity classification, access controls) in machine-readable form and register them in a policy engine

2. Pipeline integration:
Every ingestion/transformation pipeline retrieves applicable policies, enforces tagging/classification, logs lineage and creates audit artifacts.

3. Lineage & metadata automation:
As data flows across domains, the system updates lineage graphs automatically — transformations, joins, model inputs, outputs.

4. Access & usage controls:
Self-service data consumers apply for data products; policy-driven access determines whether usage is permitted, logged, throttled.

5. Audit & reporting:
Data engineering dashboards automatically surface policy- violations, data flows skipping classification, model training on un-tagged data, etc.

6. Governance feedback loop:
Governance metrics (e.g., number of policy exceptions, time to remediate un-classified data, number of audit findings) feed back into the engineering roadmap.

Key capabilities

- Policy-as-code repository (versioned, testable)
- Policy-execution graph that maps rules, lineage, and model inputs
- Automated classification & tagging of datasets, tables, columns
- Dynamic lineage graph generation with model-input mapping
- Integrated access controls tied to metadata and policies
- Governance dashboards & alerting for exceptions and audit readiness

GE Aerospace

GE Aviation implemented a self-service data platform called "SSD" (Self-Service Data) that integrated governance checks directly into data product production workflows. As part of the rollout, the data engineering team required that every dataset be tagged (classification), linked to a data owner (steward), and pass an automated schema & access-policy check before focusing downstream. Any user can look up datasets for which they have access; datasets must be tagged appropriately; once rules pass checks, the project can push to production.

Source: Dataiku

This is governance by design: pipelines retrieve policy at runtime, enforce tagging and access controls, and generate lineage/audit artifacts automatically. For leaders, the GE Aviation case shows how a governance guardrail can be built into your ingestion-to-catalog-to-production flow - turning manual gating into pipeline-enforced policy.

Insight: 2025 State of Enterprise Data Governance report

According to the Enterprise Data Strategy Board, 54% of organizations say their next-gen governance programs focus on adding governance into data workflows and increasing automation.

Thus, firms are shifting from governance as a standalone discipline, such as policy committees, toward governance as policy-as-code, and audit logs as built-in artifacts of pipelines. For teams, this means that every new pipeline, transformation, and data product must carry governance-by-default (classification, lineage, access control, audit logging) that's pre-wired before the data is used. Governance by design transforms compliance from a bottleneck into a background system that's continuous, auditable, and adaptive.

How to put these insights into action

- Treat governance as engineering work, not committee work. Build policy-as-code and automate enforcement rather than manual approvals.
- Instrument every pipeline: lineage, classification, access logs should be captured automatically.
- Measure your governance health: e.g., "percentage of datasets with classification tags", "time from dataset creation to policy application", "number of un-governed pipeline runs".
- Build a self-service layer under governance: data consumers should get fast access to datasets, but always under enforced policies and audit-ready controls.

Step 3: Optimize for cost & scale automatically

As AI workloads multiply, scale and cost control become continuous optimization problems, not quarterly budget exercises. Costs expedite as data volumes and workloads increase. Predictive autoscaling and workload-placement engines apply AI to anticipate demand spikes, reallocating compute across clusters or clouds automatically.

What a flow looks like

1. Observability & telemetry:
The data platform captures metrics on usage: which data products are accessed, by whom, how often, the compute/storage consumed per pipeline/model.

2. Usage intelligence engine:
A machine-learning or rules engine analyses trends and identifies inefficiencies like under-utilised datasets, stale pipelines, or compute spikes.

3. Predictive scaling:
Based on forecasted usage (for models, analytics, agents), the system proposes or auto-executes scaling actions: spin up/down compute, migrate workloads to cheaper tiers, archive stale data.

4. Cost attribution and chargeback:
Cost is broken down by data product/team/pipeline; teams get clear visibility into "$ per data product" and their usage patterns.

5. Feedback loop:
Engineering and finance teams use dashboards to track cost trends, reinvest savings into innovation rather than unchecked spend.

Key capabilities

- Platform-wide usage telemetry (storage, compute, network)
- AI workload analysis and anomaly detection (unexpected cost spikes, idle compute)
- Autoscaling or scheduled scaling logic tied into platform
- Cost-attribution dashboards & showback/chargeback mechanisms
- Continuous archival, tiering and lifecycle management of data assets

Real world success

General Mills reported that AI models now evaluate 5,000+ daily shipments, producing >$20M in transportation savings since FY2024 and projecting >$50M in waste reduction this year by using real-time performance data in manufacturing. The pattern is classic Step-3: platform-wide usage telemetry (shipments, lanes, dwell), a usage-intelligence engine that flags inefficiency (empty miles, service-level risk), and predictive scaling of compute to run planning models continuously instead of in nightly batches. Result: lower unit costs, faster replans, and less over-provisioned compute.

Source: CIO Dive

Maersk rolled out Trade & Tariff Studio, an AI-powered platform that centralizes customs data and reduces overpaid duties (avg. 5–6%), while cutting delays tied to poor document prep (~20% of shipments). Under the hood, this is a cost-at-scale story: centralized lineage + policy logic over fragmented customs data, intelligent workload placement for inference at peak filing windows, and predictive scaling to handle surges during regulatory changes - minimizing both compute waste and duty leakage.

Source: Maersk

How to put these insights into action

- Don't treat infrastructure cost as "someone else's problem" — include data engineering in cost-metrics and optimisation loops.
- Deploy telemetry at pipeline & model-level: storage used, compute time, number of downstream consumers. Without telemetry you cannot optimise.
- Build autoscaling and lifecycle management into your platform: e.g., archive or compress older raw data automatically; downsize compute after training finishes; schedule inference runs during off-peak windows.
- Provide transparent cost attribution: teams know the cost of their data products, are motivated to reuse and optimise rather than create redundant copies.

Step 4: Shift from management to enablement

True maturity arrives when data engineers stop managing access and start enabling creation. AI copilots, governance guards, and adaptive catalogs turn your platform into a living data ecosystem - one where every team can innovate safely, and every insight strengthens the system.

What a flow looks like

1. Data product catalog & self-service access:
The platform exposes a catalog of vetted, quality-assured datasets (data products) with clear metadata, lineage, and quality scores. Consumer teams can browse, request access and get provisions automatically where policy allows.

2. Context & guidance layer:
Embedded within the portal are usage guidelines, trust scores, recommended models or transformations, built-in governance and catalogue discovery semantics.

3. Consumption & feedback:
Business analysts, data scientists, product teams consume data products, build models/analytics, and provide feedback (ratings, issues) into the system.

4. Platform support & governance guardrails:
The data engineering team provides guardrails (security, access, cost visibility), monitors usage, and advises on reuse, not just denies access.

5. Innovation loop:
Usage data (which datasets are consumed, which aren't, time-to-insight) feeds back into the product roadmap of data engineering, enabling continuous improvement of the platform.

Key capabilities

- Catalog/search/discovery interface with lineage, metadata, trust scores
- Self-service provisioning workflows with governance embedded
- Usage analytics (which data products are popular, how long to access, how many models built)
- Embedded context: suggestions, documentation, data-science templates
- Data-engineering roadmap driven by usage insights and consumer feedback

AUTODESK

Autodesk deployed a metadata-driven, self-service enterprise data platform to serve over 13,000 employees globally. By centralizing ingestion, transformation and orchestration frameworks (via Snowflake, dbt and Fivetran), Autodesk's data engineering team replaced the legacy model of "submit a request → wait weeks" with on-demand data product access from a searchable catalog.

Key enablers included: a unified metadata layer, self-service provisioning workflows, embedded governance guardrails, and usage feedback loops. Business analysts could find, request and access vetted datasets within minutes. This case demonstrates the power of treating each dataset as a product, establishing a catalog interface, and shifting your org charter from control to enablement.

Source: Fivetran

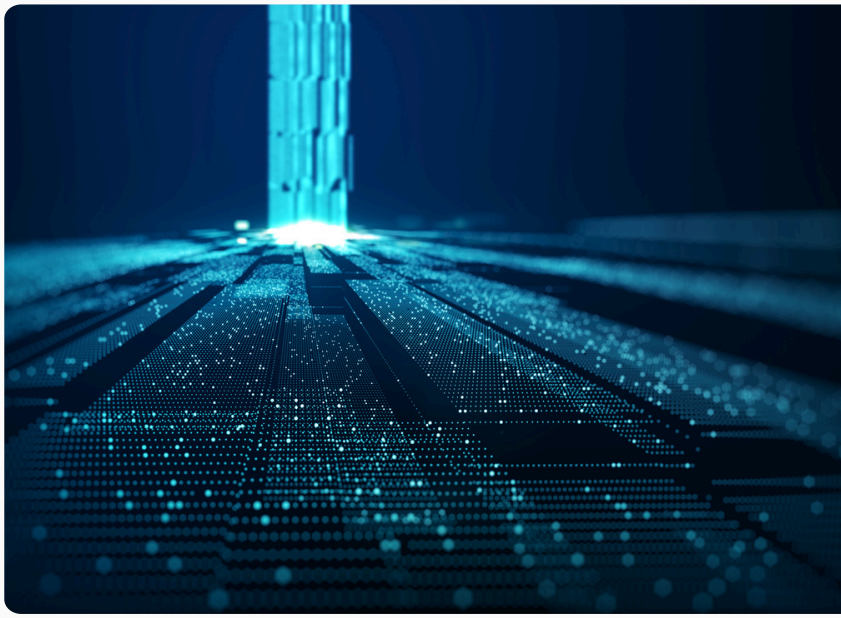
How to put these insights into action

- Build a data-product mindset: treat each dataset as a product with lifecycle, consumers, feedback, trust score.
- Empower your consumers: less "submit a request and wait 3 weeks" and more "browse, click, get access (if policy allows)".
- Track adoption: how many datasets are consumed? How many models are built by business teams? What is time-to-insight?
- Your team becomes a platform-team, not just an operational team. The organizational charter adapts.

Unframe

The AI data operating model: from platform to value

How does data engineering maturity translate to measurable AI outcomes?



Modern data engineering teams are evolving from pipeline operators to platform architects who enable the entire enterprise to innovate responsibly. The four disciplines we just explored - autonomous quality, governance by design, cost intelligence, and enablement - come together in what we call the AI Data Operating Model.

Each phase transforms a technical foundation into business value, guided by automation and AI feedback loops.

At the start, data engineers focus on centralization and trust - building a unified foundation with continuous quality monitoring. As organizations mature, governance becomes embedded as policy-as-code, turning compliance into a native feature of every pipeline. Scaling follows naturally when telemetry and machine learning drive predictive autoscaling and cost attribution. The destination is enablement—a self-service environment where data consumers can discover, build, and deploy confidently while the platform enforces guardrails in the background.

Next, **the Journey → Value Framework** captures this evolution. It shows how each discipline compounds the next, creating an autonomous, AI-driven data ecosystem that learns, optimizes, and scales with the enterprise.

The AI Data Journey → Value Framework

Journey Stage	Engineering Focus	AI-Enabled Capability	Team Shift	Business Value
Centralize → Trust	Unify datasets and metadata; instrument quality metrics	AI agents automate discovery, tagging, and data-health scoring	From finding data to trusting data	60 % faster data-to-model readiness
Govern → Assure	Encode policy as code; auto-track lineage	AI enforces access, retention, and anomaly detection in real time	From manual review to built-in compliance	Full auditability with near-zero release delay
Optimize → Scale	Deploy telemetry, predictive autoscaling, and lifecycle management	ML forecasts usage, right-sizes compute, and archives idle data	From cost control to cost intelligence	~30 % reduction in infra spend; faster training
Enable → Empower	Launch governed self-service catalog and feedback loops	GenAI copilots recommend datasets, joins, and reusable models	From gatekeeping to enablement	2× faster innovation and AI adoption

Your path forward for putting principles into practice

The transformation to AI-first data management isn't a single initiative—it's an operating model shift.

By now, the principles are clear: autonomous quality, governance by design, cost intelligence, and enablement form the foundation of every AI-ready enterprise. But success depends on translating these principles into tangible, staged action.

To help guide execution, the following roadmap outlines how you can move from concept to implementation over the next 12–18 months.

Audit your current state
What % of datasets pass automated quality checks? How many pipelines have embedded policy enforcement? What is your cost per data product? How many business teams consume your datasets in self-service mode?

Prioritize one capability per step
For the next 12–18 months, choose to build an autonomous quality engine, implement policy-as-code, enable usage telemetry and cost attribution, or launch a data product catalog.

Build the platform and team for scale
Data engineers must shift from heroic "pipeline builders" to "platform engineers" — building discoverable, governed, reusable data products for AI consumers

Measure differently
Move KPIs from number of ETL jobs, TBs ingested, or downtime, to readiness for AI, time-to-model-data, cost per data product, and number of business consumers.

Communicate the value
Use case studies (such as S&P Global's AI-ready metadata and Maersk's cost-optimisation) to show the C-suite and board members how this transformation delivers measurable results.

By following this four-step framework — autonomize quality, embed governance, optimize cost and scale, and enable teams — data engineering becomes the strategic foundation for enterprise AI. The old rules no longer suffice; these are the new rules for managing data with AI.

Unframe

Achieve AI-native data management in days

For leaders ready to operationalize this model today, Unframe acts as the intelligence layer that makes these four disciplines autonomous.

This shift requires something beyond orchestration or tooling. It calls for AI-native data management - systems that continuously tag, cleanse, optimize, and govern data autonomously.

That's what Unframe delivers. Instead of adding another tool to your stack, Unframe makes your existing data ecosystem more intelligent. Our platform deploys tailored AI workflows that automatically enhances data quality, governance, and performance - in days.

These agents can:

- Tag and cleanse data dynamically across silos and pipelines
- Detect anomalies and schema drift in real time
- Optimize data lifecycle and storage for cost and sustainability
- Discover and classify datasets automatically for compliance
- Enforce fine-grained access controls with human-in-the-loop validation
- Curate and validate data for AI readiness

The outcome is simple yet powerful: high-quality, cost-optimized, governed data - delivered through the stack you already own, made intelligent with Unframe. For leaders, this represents the next frontier in: evolving from managing data for AI to managing data with AI.

Unframe

About Unframe

Unframe is the Managed AI Delivery Platform that delivers AI solutions that work — tailored to your business and delivered in days. With Unframe, organizations can turn high-value AI use cases into real outcomes, without sharing data, committing upfront costs, or compromising on results. Unframe is headquartered in Cupertino, California, with a global presence in Tel Aviv and Berlin.

unframe.ai