

The 2026 *Manufacturing Data Capacity Gap*

How industrial leaders are closing the data
bottleneck between AI ambition and production.

AN EBOOK BY MATILLION'S MAIA



Contents

Part 1 - The Gap **03**

The Fork in the Road for Global Manufacturing 04

Three Trends Reshaping Manufacturing Data Engineering 05

Why Manufacturers Get Stuck 06

Part 2 - Three Capabilities to Close the Gap **07**

Consolidate the Stack: A Governed Execution Backbone 08

Scale Your Team Without Hiring: Agentic Data Engineering 09

Accelerate the Roadmap: Production-Ready Data in Weeks, Not Months 10

Part 3 - Proof and Path Forward **11**

Customer Spotlight: Edmund Optics 12

The Foundation for AI-Ready Manufacturing 13

Your 2026 Action Plan 14

01

The *gap*

The structural barriers preventing data-driven business outcomes.

In 2026, manufacturing is splitting in two. A small group of leaders is moving AI workloads into production. The rest are stuck behind a data engineering capacity gap. This section explains what that gap looks like, why it forms, and why hiring alone will not close it.

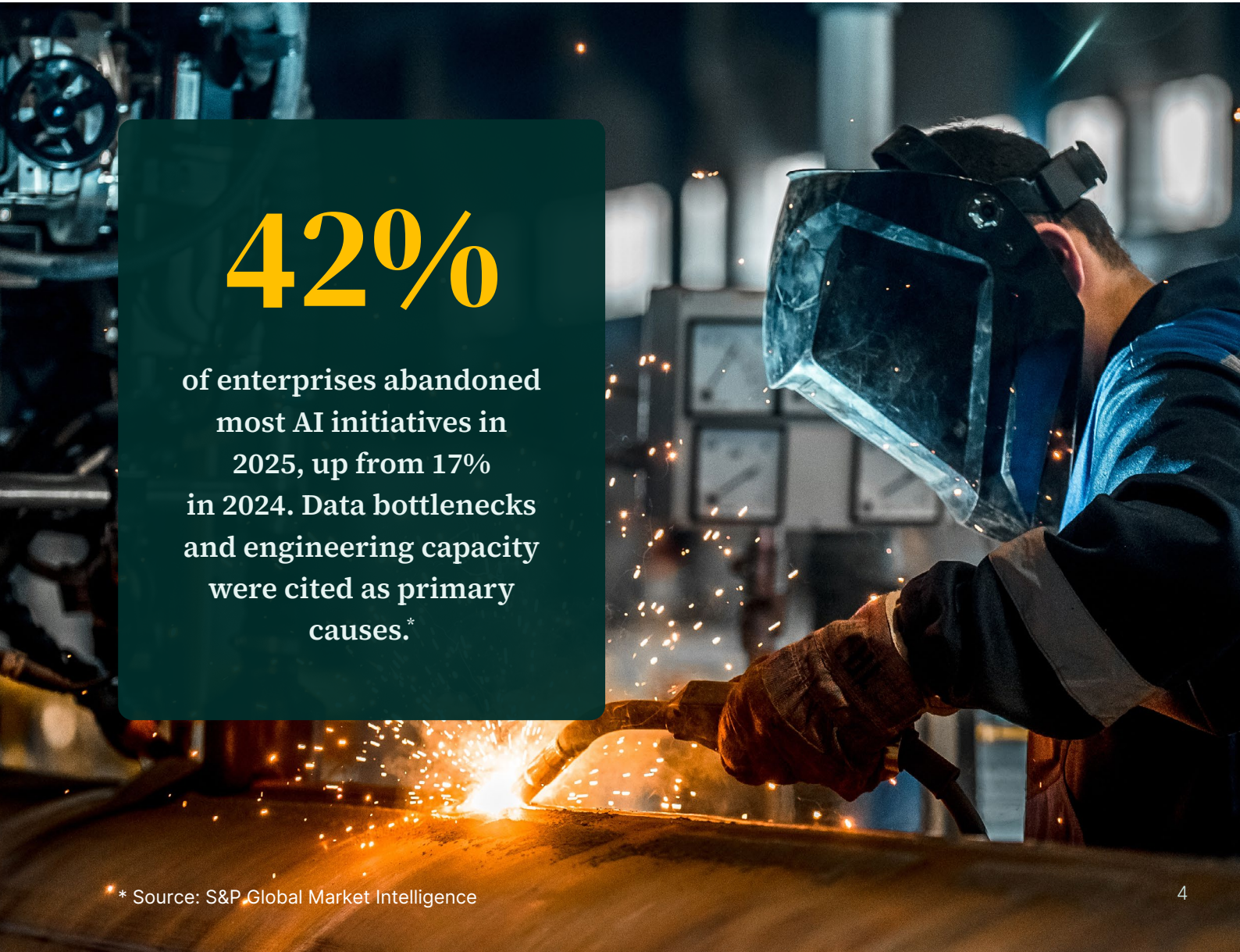
The fork in the road for *global manufacturing*

In 2026, manufacturing is splitting in two. A small group of leaders is moving AI workloads into production. The rest are stuck.

S&P Global Market Intelligence reports that 42% of enterprises abandoned most AI initiatives in 2025, up from 17% the year before. The cause is not vision. It is data. IDC finds only around 30% of organizations feel their data is ready for generative AI, and Gartner and IDC have consistently named data quality the number-one barrier to AI implementation.

The barrier is a structural Capacity Gap. Trusted, production-ready data cannot be delivered fast enough across ERP, MES, PLM, IoT, and legacy ETL to meet rising demand. Strategy is not the constraint anymore. Pipeline throughput is.

This ebook explains how manufacturers close that gap. Not by hiring endlessly, but by changing how data engineering work is produced, governed, and scaled.



42%

of enterprises abandoned most AI initiatives in 2025, up from 17% in 2024. Data bottlenecks and engineering capacity were cited as primary causes.*

* Source: S&P Global Market Intelligence

Three Trends Reshaping *Manufacturing Data Engineering*

Readiness, not experimentation, defines the next phase of manufacturing.

The question is no longer whether data demand will grow, but whether teams can keep up. **Three shifts are reshaping the landscape.**

01 AI is moving from analysis to autonomous operational execution

Manufacturers are no longer using AI only to explain what happened. AI now sits inside day-to-day workflows: production planning, predictive quality, condition-based maintenance, and operational decision-making in real time.

That raises the bar for the data layer. Reporting after the fact is no longer enough. Manufacturers need production-ready pipelines delivering data on a timeline the business can act on, with lineage and governance enforced at build time.

02 Fragmented data estates are getting more expensive to run

What used to be tolerated as architecture complexity is now a direct barrier to AI adoption. ERP systems (SAP, Oracle, Microsoft Dynamics), MES, PLM, IoT, and legacy ETL platforms such as Informatica PowerCenter, Alteryx, SSIS, and IBM DataStage are spread across plants, business units, and acquisitions. Teams stitch pipelines together by hand.

The result: slower delivery, weaker governance, and a data estate that cannot support AI at scale.

03 Human-only delivery models are hitting their limit

Industry surveys show 80% of data engineers struggle to keep up with business demand. The 2024 CDAO Agenda found 70% of data and analytics leaders are now expected to deliver faster with fewer resources.

Senior engineers spend their time fixing, documenting, and maintaining instead of building. Delivery slows, hiring gets harder, and AI programs stall. The teams that move fastest in 2026 will expand delivery capacity without scaling headcount linearly.

Why manufacturers *get stuck*

Five structural realities drive pilot stagnation in industrial environments.



ERP Sprawl

Multiple, fragmented ERP systems across plants and acquired subsidiaries. Each site accumulates its own databases, decentralizing critical data. Leadership lacks a holistic view, so simple tasks like consolidated inventory, Days Inventory Outstanding, or total product line cost are slow, inconsistent, or both.



Unstructured Data Overload

Factories generate equipment logs, engineering PDFs, safety checklists, inspection images, and other operational records. Traditional ETL tools struggle to ingest, govern, and operationalize these formats inside a governed pipeline.



Contractor Dependence

Lean data teams supplement analytics with external SI and consulting partners. Contracting fills the queue but does not build leverage. When the contract ends, the pipelines remain, the documentation is thin, and the institutional context walks out with the consultant.



The AI Readiness Gap

Gartner and IDC name data quality the number-one barrier to AI implementation. Only around 30% of organizations feel their data is AI-ready. Pilot projects languish in sandbox environments because moving them to production exposes pipeline issues legacy infrastructure cannot solve.



Legacy Stack Complexity

Industrial estates run a generation of MES, PLM, and on-premise ETL platforms that were never designed for cloud-native AI workloads. The data is siloed away from modern analytics. Even small upstream changes (a schema rename, a connector update) disrupt downstream delivery.

02

Three capabilities *to close the gap*

Consolidate the stack.

Scale your team without hiring.

Accelerate the roadmap.

Closing the manufacturing data capacity gap requires more than faster ETL. It requires a governed automation layer that can consolidate fragmented tools, expand engineering capacity, and turn AI data readiness into a repeatable operating model. Maia by Matillion is an AI Data Automation platform built for that shift. It combines autonomous data engineering agents, organizational context, and an enterprise-grade data platform foundation to help teams build, migrate, maintain, and govern production pipelines at scale.

Consolidate the Stack: *A Governed Execution Backbone*

Most manufacturers run a stack that looks like this: Informatica or SSIS for ingestion, Alteryx for transformation, hand-built Airflow DAGs for orchestration, and dashboards stitched on top for monitoring. Every tool has its own pricing, its own governance model, and its own learning curve.

The first step is consolidating fragmented data workflows into a governed execution backbone. For manufacturers, that means bringing ingestion, transformation, orchestration, lineage, and operational control into a single cloud-native platform rather than managing them across separate tools.

Autonomous Conversion.

Maia's legacy migration agent converts pipelines from platforms such as Informatica PowerCenter, Alteryx, IBM DataStage, SSIS, Oracle ODI, SAS Enterprise Guide, dbt, and

others into cloud-native logic, with no manual rewrite and no GSI dependency. In a public webinar on March 31, 2026, the platform converted 100 Informatica PowerCenter pipelines in 30 minutes, with output rebuilt natively for Snowflake, Databricks, or Amazon Redshift.

Unified Backbone.

Harmonize ERP, MES, and PLM data into a single operational lens in Snowflake, Databricks, BigQuery, or Redshift. This eliminates the dual-governance overhead and fragility of running data outside the warehouse.

Lineage by Default.

Every job emits lineage and naming-standard metadata at build time. Audit trails are enforced during creation, not chased down quarterly. Compliance becomes a byproduct of the work.

PROOF POINT

UK energy distribution leader:

Separating from a parent group and building an independent data platform from scratch, the company faced 100+ complex Alteryx workflows. The platform migrated all of them into Snowflake in five weeks, eliminating \$300K+ per year in licensing for legacy data tooling.

100+

Alteryx workflows
migrated

5 wks

from scratch to
production-ready

\$300K

annual license
spend eliminated

Scale Your Team Without Hiring: *Agentic Data Engineering*

Manufacturers need to expand delivery capacity without scaling headcount linearly. The winners in 2026 are not the companies that hire the most engineers. They are the ones whose engineers ship the most. Agentic data engineers give your existing team that leverage.

Eliminate Toil.

Across recorded POVs and production deployments, agentic data engineering has reduced manual effort by 90–95% in common workflows such as troubleshooting failed jobs, updating legacy transformations, and generating documentation. That capacity goes back to the team, so engineers can focus on higher-value work.

End-to-End Execution.

Agentic data engineers understand business intent and execute end-to-end pipelines inside a governed environment. This is not a copilot suggesting code snippets in an IDE. It is a teammate shipping pipelines into production under the same lineage, validation, and approval gates a human engineer would use.

Capture Knowledge.

The Context Engine encodes naming standards, lineage requirements, and unique business rules so institutional know-how stays inside the platform, not in the heads of contractors who eventually leave.

PROOF POINT

Balfour Beatty

Balfour Beatty

FTSE-listed infrastructure firm building nuclear power stations and high-speed rail. The 40-person data team was spending roughly 80% of its time on routine pipeline maintenance, leaving 20% for the predictive analytics that matter on safety-critical projects.

After deploying agentic data engineering, standard pipeline delivery dropped from three weeks to three hours, and Informatica PowerCenter XML analysis fell from one week to six minutes.

“The technology gives the team extra capacity for heavy lifting. It does not take away the role of the data team; it enhances what the team can deliver.” - Mark Hume, Head of Data, Balfour Beatty

Accelerate the Roadmap:

Production-Ready Data in Weeks, Not Months

To escape pilot purgatory, manufacturers need a repeatable path from idea to production-grade data. AI Data Automation accelerates the roadmap by automating the work of preparing, moving, and transforming data.

10x

Faster data delivery

Complex pipelines and migrations that once took weeks now take hours. Across the customer base, teams typically see a 10× improvement in throughput per engineer. In headline cases, customers have reported up to 100× gains after replacing legacy tools.

35% / 60%

Faster, Validated Migrations

At GE HealthCare, automated migration of RSQL scripts and stored procedures into Snowflake delivered 35% faster pipeline development and 60% faster validation. This accelerated access to AI-ready clinical and operational data.



GE HealthCare

5d > 30m

Documentation at Machine Speed

At Sophos, automated documentation and testing workflows compressed Jira-based work from five days to 30 minutes. That is 80× faster, with documentation reverse-engineered and regenerated automatically.

SOPHOS



Governed Speed

Automating lineage and documentation lets manufacturers accelerate delivery across R&D, quality, and production without trading off governance. Compliance is enforced at build time, not reconstructed at audit time.

03

Proof and *Path Forward*

How Edmund Optics scaled without hiring,
the foundation for AI-powered manufacturing,
and a 2026 action plan.

The organizations that win in 2026 will be those that turn demand into governed execution. The remaining chapters show what that looks like in practice. A customer who avoided five hires. The structural shifts that make it possible. A four-step plan to get there.

CUSTOMER SPOTLIGHT



Edmund Optics

Scaling Data Strategy Without the Hiring Spike.

For global manufacturer Edmund Optics, the challenge was scaling a complex data strategy with a lean internal team.

The challenge

Rising demand for data across the organization projected a need for five additional data engineering hires, which would have significantly increased operating overhead.

The transformation

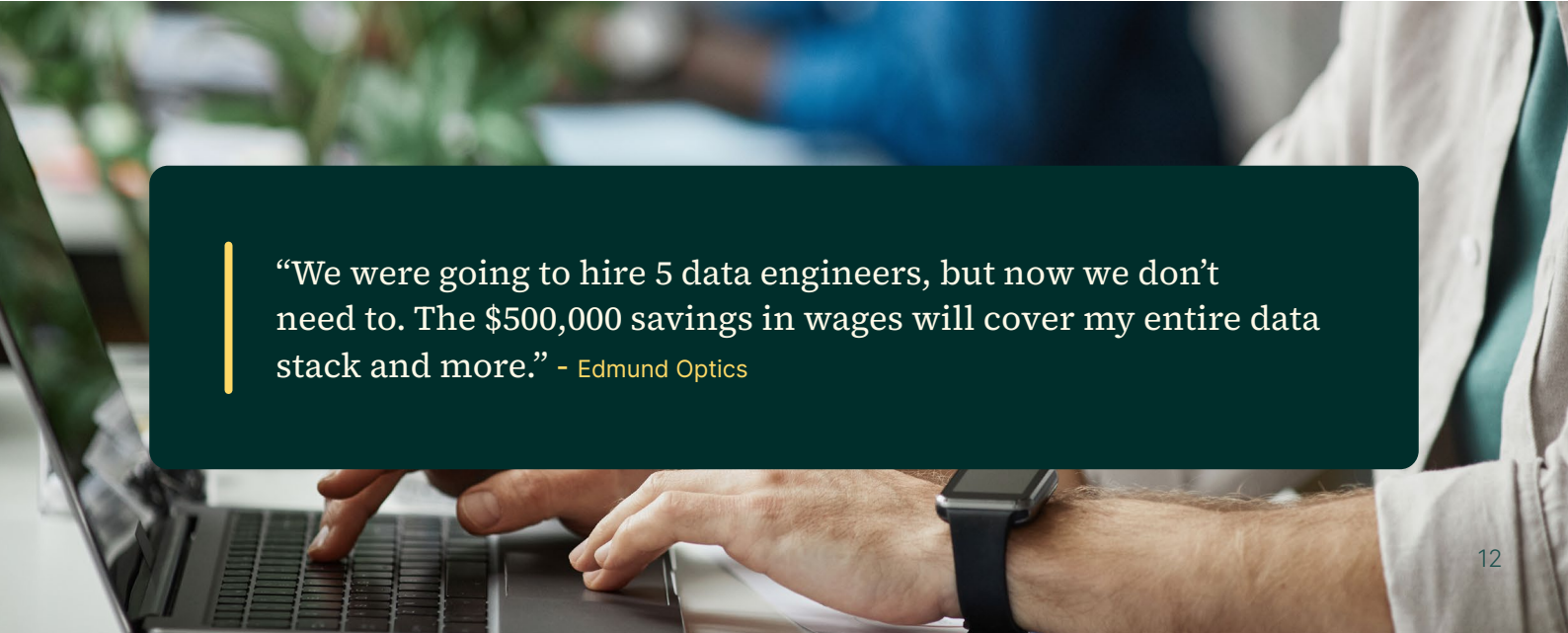
Instead of adding fixed headcount, they used AI Data Automation to augment existing capacity and empower their existing staff to move faster.

The impact

Empowering a single junior engineer with the platform allowed the company to successfully avoid all five new hires.

The outcome

The savings in wages alone, estimated at \$500,000, covered the cost of their entire data stack while enabling a faster self-serve environment.

A close-up photograph of a person's hands typing on a laptop keyboard. The person is wearing a white shirt and a black smartwatch. The background is blurred, showing an office environment with plants and other people.

“We were going to hire 5 data engineers, but now we don’t need to. The \$500,000 savings in wages will cover my entire data stack and more.” - Edmund Optics

The Foundation for *AI-Ready Manufacturing*

The organizations that win in 2026 will be those that turn demand into governed execution. AI Data Automation provides the foundation by replacing hand-crafted, human-only delivery with three structural shifts.



One runtime. Multiple legacy tools retired.

Ingestion, transformation, orchestration, and lineage in a single cloud-native platform. Replaces Informatica, Alteryx, and SSIS, and removes the dual-governance overhead of running data outside the warehouse.



Agentic data engineers, not copilots.

Agentic data engineers ship end-to-end pipelines from business intent inside a governed environment. Not a copilot suggesting code snippets, but a teammate shipping production code under the same controls a human engineer would use.



Your company's way, encoded.

The Context Engine captures naming standards, lineage requirements, and business rules so the platform builds your company's way from day one. Institutional knowledge stays inside the platform when people leave.

“Maia offers a glimpse into the future of data engineering. It’s intuitive, powerful, and feels like a real accelerant for how teams build with data.” - Sridhar Ramaswamy, CEO, Snowflake

Your 2026 *Action Plan*

Success in 2026 does not come from trying to deploy AI everywhere at once. It starts by using AI to fix the data engineering bottleneck.

01

Consolidate the stack

Replace Informatica, Alteryx, and SSIS with a single governed backbone. Cut technical debt, software costs, and infrastructure overhead.

02

Scale the team without hiring

Deploy agentic data engineers to handle repetitive toil: pipeline maintenance, documentation, and troubleshooting. Your engineers focus on production AI delivery.

03

Accelerate production delivery

Shorten the path from pilot to production by removing the plumbing delays and manual integration hurdles that kill project momentum.

04

Prioritize high-value workflows

Start with a small set of high-impact use cases such as predictive maintenance, consolidated inventory, and supplier risk. Prove the model and scale beyond pilot purgatory.

Ready to solve the data engineering capacity constraint?

Book your data readiness assessment today.

[Book your assessment](#)



An AI Data Automation platform built to close
the manufacturing data capacity gap.

Maia by Matillion combines autonomous data engineering agents, an enterprise Context Engine, and a cloud-native execution backbone. Manufacturers consolidate fragmented stacks, scale delivery without scaling headcount, and move AI workloads from pilot to production.

[Book your assessment](#)

© 2026 Matillion Limited. All rights reserved.

Maia and the Maia logo are trademarks of Matillion Limited. All other product, brand, and company names referenced herein are trademarks of their respective owners and are used for identification purposes only.

Your 2026 *Action Plan*

Success in 2026 does not come from trying to deploy AI everywhere at once. It starts by using AI to fix the data engineering bottleneck.

01 Consolidate the stack.

Replace Informatica, Alteryx, and SSIS with a single governed backbone. Cut technical debt, software costs, and infrastructure overhead.

02 Scale the team without hiring.

Deploy agentic data engineers to handle repetitive toil: pipeline maintenance, documentation, and troubleshooting. Your engineers focus on production AI delivery.

03 Accelerate Production Delivery.

Shorten the path from pilot to production by removing the plumbing delays and manual integration hurdles that kill project momentum.

04 Prioritize High-Value Workflows.

Start with a small set of high-impact use cases such as predictive maintenance, consolidated inventory, and supplier risk. Prove the model and scale beyond pilot purgatory.

Ready to solve the data engineering capacity constraint?
Book your data readiness assessment today.

[Book your assessment](#)