

# AI Success Starts with Data

---

A Strategic Guide 

## FOREWORD

# AI Success Hinges on the Strength of Your Data

Successful Artificial Intelligence (AI) adoption requires a solid foundation, and that starts with your data.

We have all seen the reports highlighting a staggeringly high failure to launch when it comes to Generative AI initiatives. About 30% of pilot efforts stall out.<sup>2</sup> Industry research shows the biggest obstacle for 45% of companies is issues with fragmented, unstructured data.<sup>1</sup>

The impacts are real, leading to 87% of initiatives never making it to production.<sup>3</sup> Even worse, Massachusetts Institute of Technology (MIT) found that of the efforts that do deploy, only 5% result in meaningful improvement<sup>4,5</sup> — these organizations had solved their data problems.

It's become painfully clear that tackling problems at the source is more critical than ever.

For these reasons, Resource Data deploys cross-functional teams with expertise in Data Science & Engineering, Analysis, Systems Engineering, Software Development, and AI to help companies like yours ensure that AI initiatives have a solid foundation to build on.



**Jason Child**  
AI Service Area  
Lead



**Rick Busch**  
Data Engineering  
Service Area Lead

# Contents



- 0 Executive Summary
- 1 The Costs of Poor Data
- 2 A Roadmap to AI-Ready Data
- 3 AI Data Readiness Checklist & Action Plan
- 4 Emerging Trends
- 5 How Resource Data Can Help
- 6 The Destination

# EXECUTIVE SUMMARY

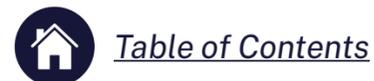
## AI Success Depends on Data Readiness

Up to 87% of AI projects fail to reach production,<sup>3</sup> often due to poor, fragmented, or inconsistent data. Even surviving projects face long delays, with eight months on average spent preparing data instead of innovating. The costs are steep: \$12.9 million in annual losses per organization,<sup>6</sup> wasted expert time,<sup>7</sup> and reputational damage from unreliable AI outputs.<sup>8,9</sup>

## Data readiness is a strategic imperative

A disciplined approach — Assess, Remediate, Govern — transforms scattered, low-quality data into a trusted, scalable foundation for AI. Leaders who treat data as a strategic asset can accelerate AI ROI, reduce risk, and future-proof their organizations in an AI-driven market.

Resource Data | AI Success Starts with Data: A Strategic Guide



[Table of Contents](#)

“ —  
Without a strong data foundation, AI is a gamble. With it, AI becomes a predictable engine for value.

## What You'll Gain from This Report

- A **proven three-phase roadmap** to achieve AI-ready data
- An **executive checklist** to quickly assess your organization's AI readiness
- Real-world **case study demonstrating ROI** from disciplined data preparation
- Insights on **emerging trends** — AI-driven automation, synthetic data, and real-time observability — that can future-proof your AI strategy

# 01

## The Costs of Poor Data

---

Neglecting data preparation is not a minor oversight; it is the leading cause of AI project failure.

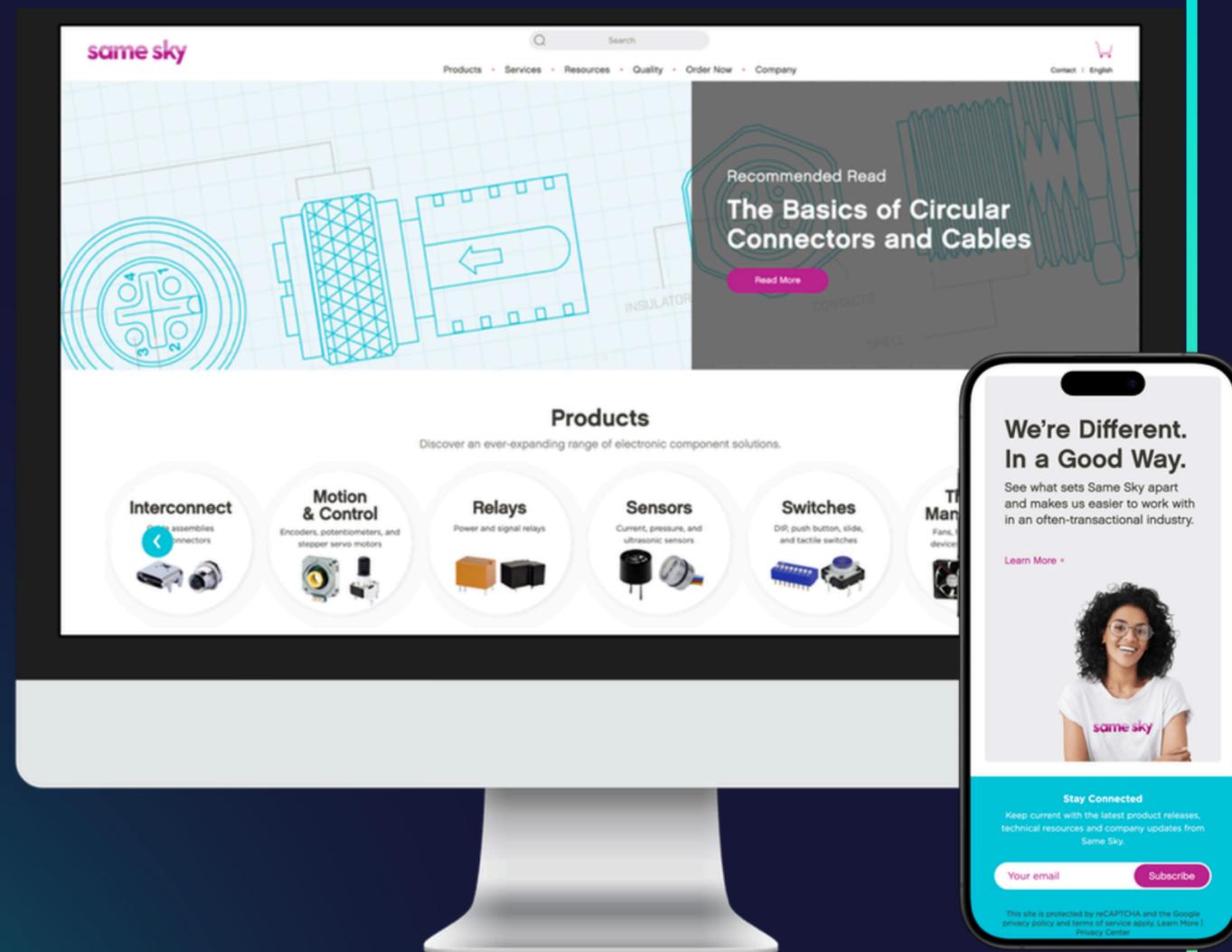
# Understanding the Cost of Poor Data

Data debt silently taxes all phases of AI initiatives. Beyond finance, missteps with an externally facing AI solution can torpedo customer trust and brand equity.

The financial, operational, talent, and reputational costs are substantial and well-documented.

| Impact Domain      | Consequences                             | Research Highlights   |
|--------------------|--|---|
| Project Viability  | Pilots stall; Scale fails                | ≤48% make production; up to 87% fail <sup>3</sup>                                   |
| Timelines          | Extended project timelines               | Average 8 months from POC to production, primarily data prep <sup>2</sup>           |
| Talent Misuse      | Experts diverted to data custodial tasks | Data scientists spend 50-80% on data manipulation and cleaning <sup>7</sup>         |
| Financials         | Direct losses and wastage                | \$12.9 million/year <sup>2</sup> ; \$3.1 trillion economy-wide losses <sup>10</sup> |
| Trust and Adoption | Users abandon unreliable tools           | 84% of customers churn after data-driven errors <sup>11</sup> .                     |

# Case Study



## RAG Chatbot Improves Customer Service

A global electronics manufacturer and distributor wanted to use AI to transform engineering support and product development, but fragmented, inconsistent data stood in the way. Rather than jumping into predefined AI use cases, Resource Data first profiled their datasets to expose gaps that would have derailed progress.

Through a disciplined assess, remediate, and govern approach, we converted raw data into a reliable foundation, enabling the iterative development of retrieval-augmented generation (RAG) pipelines. The chatbot helps engineers quickly find and synthesize precise technical information, accelerating decision-making and innovation.

# 02

## A Roadmap to AI-Ready Data

---

Build the bridge between AI ambition and on-the-ground data reality.

# AI-Ready Data: A DISCIPLINED ROADMAP

We understand the chasm between executive AI ambition and the on-the-ground data reality because we have helped numerous technical leaders like you build the bridge between them. The solution is not a magic bullet but a disciplined, systematic approach. An effective path to data readiness can be distilled into three core phases:

## 01

### Assess & Strategize

Clarify business goals, inventory data sources, and profile data quality and volume comprehensively.

## 02

### Remediate & Build

Clean and validate data systematically. Engineer essential features. Build AI-ready architecture.

## 03

### Govern & Iterate

Implement continuous validation and feedback loops to maintain high standards and mitigate risk.

Resource Data | AI Success Starts with Data: A Strategic Guide



[Table of Contents](#)

# 01

# ASSESS & STRATEGIZE

**You cannot fix what you do not understand. The first phase is a rigorous assessment to establish a clear baseline of your current data landscape.**

## Define AI Objectives

Begin with the end in mind. Before diving into data, first clarify the business outcomes you expect AI to achieve. This means defining the problem to solve, the decisions to improve, and the KPIs that will signal success.

Establishing these objectives keeps teams aligned, prevents “AI for AI’s sake,” and ensures that subsequent data readiness work is tied directly to measurable business value. By beginning with the end in mind, leaders create a roadmap where every data and technology decision supports strategic impact.

## Inventory and Profile Data

Identify and inventory all relevant data sources, from structured databases to unstructured PDFs and logs. This is the time to break down data silos and consolidate access to create a unified view.

Use data profiling tools to analyze the structure, uncover initial quality issues like missing values and inconsistencies, and understand the true state of your assets. This initial discovery process answers the first critical question: “What usable data do we actually have?”

# 02

# REMEDiate & BUILD

With a clear understanding of your data's condition, the work of transformation begins. This phase focuses on systematically improving data quality and building the infrastructure to support AI workloads.

## Execute Data Cleaning and Transformation

This process includes correcting errors, filling in missing values, removing duplicate entries, and ensuring consistent formatting. When preparing data for AI models, it also means normalizing numerical data and converting categorical information into machine-friendly formats.

## Engineer the Right Features

Raw data is insufficient. Feature engineering creates useful features from existing data, enhancing AI model performance. This includes structuring data for applications like RAG systems that connect language models to enterprise knowledge.

## Build an AI-Ready Architecture

Conventional data architectures lack flexibility for AI, which requires scalable data lakes or lakehouses to manage varied data types. Automated ETL/ELT pipelines are designed and implemented to deliver reliable, repeatable flows of quality data from sources to AI models.

# 03

# GOVERN & ITERATE

Data preparation is not a one-off project; it is a continuous lifecycle integrated with your MLOps practices. A “prepare once, use many times” approach is insufficient in a world where data constantly drifts and business needs evolve.

## Implement Continuous Validation

Establish automated data quality monitoring to detect issues like data drift or schema changes in real-time. Ongoing validation ensures that the quality of your data does not degrade over time, which protects the reliability of your AI systems.

## Establish Robust Governance

Formal data governance is the cornerstone of sustained AI success. This means establishing clear ownership, policies, and standards for data quality, security, and ethical use. Frameworks like the NIST AI Risk Management Framework (AI RMF)<sup>12</sup> and the ISO/IEC 42001 standard<sup>13</sup> provide guidelines for creating accountable, transparent, and fair AI systems.

## Create Feedback Loops

The AI lifecycle is inherently iterative. Establish mechanisms to collect feedback on model performance and, particularly for RAG systems, on the relevance of retrieved information. Use this feedback to continuously refine data sources and preparation processes.

# 03

## AI Data Readiness Checklist & Action Plan

---

Data must be accurate, timely, relevant, consistent, and secure.

# AI Data Readiness CHECKLIST

## ● **Assessment**

Have you defined clear AI goals and inventoried all relevant data sources?

## ● **Data Quality**

Is there a plan to handle missing values, correct errors, and remove duplicates?

## ● **Transformation**

Are you standardizing formats, normalizing features, and engineering new variables?

## ● **Infrastructure**

Is your architecture scalable? Are data pipelines automated?

## ● **Governance**

Have you established data quality monitoring and clear governance policies?

## ● **Iteration**

Is there a feedback loop to continuously improve data based on model performance?

# AI Data Readiness

## ACTION PLAN



# 04

## Emerging Trends

---

Seven data readiness shifts reshaping AI's speed, trust, and ROI.

# Data Readiness SHIFTS

1

## AI-Driven Automation

*Automates cleaning, labeling, and monitoring.*

**Why it matters:** Reduces data prep timelines from months to days, accelerating speed-to-market for AI initiatives and freeing talent for higher-value work.

2

## Synthetic Data

*Generates realistic datasets to fill gaps, reduce bias, and protect privacy.*

**Why it matters:** Enables innovation in data-scarce or highly regulated industries, letting you train models without waiting for real-world data or risking non-compliance.

3

## Data-Centric AI

*Prioritizes continuous data quality improvement over tweaking model.*

**Why it matters:** Directly improves AI accuracy and reliability. Avoids costly rework and poor model performance caused by flawed inputs.

4

## Real-time Observability

*Continuously monitors freshness, lineage, and anomalies.*

**Why it matters:** Prevents costly business disruptions by detecting data drift or quality drops before they degrade AI decision-making.



# Data Readiness **SHIFTS**

5

## AI-Agents for Data Ops

*Autonomous agents that profile, remediate, and govern data pipelines.*

**Why it matters:** Reduces operational risk and cost by automatically enforcing governance policies and optimizing flows without human bottlenecks.

6

## Multimodal Data-Readiness

*Prepares text, image, audio, and video for integrated AI systems.*

**Why it matters:** Unlocks richer insights and new revenue streams by enabling AI to process and reason across multiple content types.

7

## Adaptive and Personalized Retrieval

*Advanced RAG methods that tailor information retrieval to user context.*

**Why it matters:** Improves decision quality and user adoption by ensuring AI delivers the most relevant, context-aware answers every time.

# EMERGING TRENDS

## Key Takeaways

Leaders who monitor AI developments and selectively invest in them will shorten AI deployment cycles, safeguard compliance, and unlock competitive advantages that others will struggle to match.

The next wave of market leaders will not be defined by the algorithms they choose, but by the speed, precision, and trustworthiness of the data pipelines they build.

“

**These trends are not distant possibilities—they are active disruptors of how organizations prepare, govern, and leverage data for AI.**

# 05

## How Resource Data Can Help

---

Build your future on a foundation of solid engineering discipline.

# Resource Data AI SERVICES

## ● **AI and ML Operations**

Designing and implementing the infrastructure, pipelines, and operational frameworks that enable scalable, reliable, and efficient AI-driven solutions

## ● **AI Strategy**

Partnering with organizations to pinpoint high-impact AI opportunities, build internal expertise, and drive adoption through clear roadmaps and change-management initiatives

## ● **AI and Data System Readiness**

Assessing current data assets, systems, and processes to identify gaps, mitigate risks, and establish the foundational conditions required for successful AI and machine learning (ML) initiatives

## ● **AI Security**

Embedding security, compliance, and fairness into AI and ML solutions to safeguard against threats, protect sensitive data, and minimize bias in decision-making

## ● **Agent and ML Development**

Creating tailored, domain-specific AI agents and ML models that address unique business challenges and deliver measurable results

# 06

## The Destination

---

Successful AI shouldn't be a gamble. It should be the outcome of a deliberate, measurable, and reliable process.

# THE DESTINATION

Embarking on this disciplined journey transforms AI from a high-risk gamble into a predictable engine for value. When you build on a solid data foundation, the outcomes change dramatically.

The evolution of AI toward more advanced agentic systems and real-time applications will only intensify the need for this foundational data discipline. By mastering data readiness now, you are not just solving today's challenges; you are building the capacity to lead your organization into the future.

If you are ready to move from AI ambition to a concrete, data-led execution plan, let's talk about your data readiness roadmap. To begin your AI data journey, contact us for a comprehensive data readiness assessment.

Explore a new path.  
Focus on what matters most.

[Read Insights](#)

[Book a Meeting](#)



# References

1. Huble. *Poor data blocks AI decisions for 69% of companies. Here's why.* Accessed May 30, 2025. <https://huble.com/blog/ai-hidden-data-crisis>
2. Gartner. *Gartner Predicts 30% of Generative AI Projects Will Be Abandoned After Proof of Concept By End of 2025.* Published July 29, 2024. Accessed June 9, 2025. <https://www.gartner.com/en/newsroom/press-releases/2024-07-29-gartner-predicts-30-percent-of-generative-ai-projects-will-be-abandoned-after-proof-of-concept-by-end-of-2025>
3. VentureBeat. *Why do 87% of data science projects never make it into production?* Accessed June 9, 2025. <https://venturebeat.com/ai/why-do-87-of-data-science-projects-never-make-it-into-production/>
4. *The GenAI Divide: State of AI in Business 2025.* MIT Project NANDA (July 2025). PDF report. [https://mlq.ai/media/quarterly\\_decks/v0.1\\_State\\_of\\_AI\\_in\\_Business\\_2025\\_Report.pdf](https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf)
5. Catmull, Jaime. *MIT Says 95% Of Enterprise AI Fails — Here's What The 5% Are Doing Right.* Forbes, August 22, 2025.
6. Gartner. *Gartner Survey Finds Generative AI Is Now the Most Frequently Deployed AI Solution in Organizations.* May 7, 2024. <https://www.gartner.com/en/newsroom/press-releases/2024-05-07-gartner-survey-finds-generative-ai-is-now-the-most-frequently-deployed-ai-solution-in-organizations>
7. Informatica. *What Is Data Preparation?* Accessed June 9, 2025. (includes expert time stat: 50–80% spent cleaning/structuring data). <https://www.informatica.com/resources/articles/what-is-data-preparation.html>
8. Forbes. *IBM Watson Health's Challenges Tell Us More About Healthcare Data Than It Does About AI.* May 3, 2022. <https://www.forbes.com/councils/forbestechcouncil/2022/05/03/ibm-watson-healths-challenges-tell-us-more-about-healthcare-data-than-it-does-about-ai/>
9. Forbes. *Unity Stock: Priced Too Low For The Long-Term Opportunity.* May 20, 2022. <https://www.forbes.com/sites/bethkindig/2022/05/20/unity-stock-priced-too-low-for-the-long-term-opportunity/>
10. Harvard Business Review. *Bad Data Costs the U.S. \$3 Trillion Per Year.* September 2016. <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>
11. Grax. *Using Customer Service Analytics to Reduce Churn and Costs.* Accessed June 9, 2025. <https://www.grax.com/blog/using-customer-service-analytics-to-reduce-churn-and-costs/>
12. NIST. *AI Risk Management Framework (AI RMF 1.0).* 2023. <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf>
13. EY. *ISO 42001: Paving the way for ethical AI.* January 17, 2025. [https://www.ey.com/en\\_us/insights/ai/iso-42001-paving-the-way-for-ethical-ai](https://www.ey.com/en_us/insights/ai/iso-42001-paving-the-way-for-ethical-ai)