

RESEARCH

Why 60% of AI Projects Fail (And the Boring Business That's Cleaning Up)

Bad data costs enterprises \$3.1T a year. The companies fixing it look nothing like AI startups — and they're the most defensible bet on the table.

By Oliver Signorini · 2026-05-05

Why 60% of AI Projects Fail (And the Boring Business That's Cleaning Up)

Bad data costs enterprises \$3.1T a year. The companies fixing it look nothing like AI startups — and they're the most defensible bet on the table.

Why 60% of AI Projects Fail (And the Boring Business That's Cleaning Up)

Gartner says 60% of AI projects will be abandoned by the end of 2026. Not because the models don't work. Because the data underneath them is broken.

This is the most underpriced opportunity in business right now: the gap between *every company wants AI* and *almost no company has the data to use it*.

Here's the case, with the numbers, and what to actually build.

The pattern: boring businesses always win

The most durable wealth-generating businesses of the last 50 years share a structural pattern. They solve unglamorous, mandatory problems. They embed themselves in operational workflows. They compound through low churn.

Company	Revenue	Margin	What they do
ADP	\$18.8B	19% net	Payroll for 1-in-6 US workers
Waste Management	\$20.4B	28% EBITDA	Landfill monopoly
Cintas	\$9.8B	20% net	Uniform rental + safety
Thomson Reuters	\$6.8B	38% EBITDA	Legal/compliance workflow
Snowflake	\$3.6B	29% YoY growth	Cloud data warehouse

Look at what they have in common – not what they do, but how they hold their position:

- 1. Regulatory mandate as demand floor.** Payroll, waste disposal, compliance – demand doesn't evaporate in recessions. ADP and Paychex grew through 2008-09.
- 2. Workflow embedding.** Westlaw is open on a lawyer's second monitor. ADP is wired into a company's bank, tax IDs, and benefits providers. The switching cost isn't price – it's operational disruption, retraining, compliance liability.
- 3. Data network effects.** C.H. Robinson's 40 years of freight pricing. Dun & Bradstreet's business credit data. Invisible to customers. Impossible to replicate.
- 4. Revenue-per-customer expansion.** Cintas started with uniforms, added first aid, fire safety, document destruction. Each add-on deepens integration.
- 5. Low churn as compounding engine.** Cintas reports ~90% annual retention. At that rate, a 2000 cohort of 1,000 customers still has ~350 of those same customers in 2024.

The lesson isn't "buy a laundromat." It's that businesses sitting on regulatory mandate + workflow embedding + long relationship duration are among the most durable structures in capitalism.

The new version of this is data infrastructure.

The enterprise data mess: how bad it actually is

Before getting to the opportunity, sit with the numbers. They're worse than you think.

- The average enterprise manages **305 SaaS applications**. Large enterprises average **473**.
- **52.7% of purchased SaaS licenses go unused** — about \$21M in annual license waste per organization.
- **55-60% of enterprise data is "dark"** — collected but untagged, siloed, or never analyzed. Veritas puts functionally useless data at **85%**.
- **71% of enterprise applications remain unintegrated** — unchanged for three consecutive years despite massive iPaaS spending.
- **Only 2% of IT leaders** report integrating more than half their applications.

What it costs:

- Poor data quality costs the average enterprise **\$12.9M-\$15M annually** (Gartner).
- Across the US economy: **\$3.1 trillion per year** in losses (IBM/Gartner).
- Employees spend up to **27% of their time correcting bad data**.
- Companies running separate CRM, ERP, and supply chain systems experience **20-30% revenue leakage** from siloed inefficiencies.

A typical mid-market company (\$50M-\$500M revenue) runs 5-8 distinct business systems and has no centralized analytical layer. Their IT department

is 2-5 people. Not a data team. Not even close.

The AI readiness gap is the forcing function

This is what changes the calculus in 2026:

- **63% of organizations** don't have the right data management practices to support AI.
- **60% of AI projects will be abandoned through 2026** because orgs lack AI-ready data.
- **92% of firms plan to increase AI budgets** — while simultaneously being unable to access, clean, or trust the data those AI systems need.
- Fewer than **10% have a clear AI roadmap** with prioritized use cases.

Every company has a board pushing them to “do something with AI.” Almost none of them can. The disconnect is the opportunity.

Why AI chatbot startups are losing

Most niche AI startups are architectural wrappers: call OpenAI or Anthropic's API, apply a thin layer of prompting and UI, charge a subscription. Three structural problems make this fragile:

- 1. Margin compression.** API costs are dominant COGS. When GPT-4 pricing dropped ~80%, competitive pressure forced wrappers to pass the savings on. Gross margins sit at 40-60% versus 70-85% for traditional SaaS.
- 2. Commoditization.** When the underlying capability is available to every competitor at the same price, differentiation collapses to brand and UX — neither durable. Any competitor can copy your prompt template in a weekend.

3. Platform risk. OpenAI, Anthropic, and Google are product companies, not neutral infrastructure. When your feature becomes their feature, you lose.

The case studies:

- **Jasper AI** raised \$125M at a \$1.5B valuation. By late 2023: layoffs, missed targets, valuation markdowns. GPT-4 made their templates look marginal.
- [Character.ai](#) hit ~\$1B ARR but couldn't monetize depth. Google licensed their tech for \$2.7B — a talent acqui-hire that signalled the standalone product had limited upside.
- **Inflection AI (Pi)** raised \$1.3B, then was effectively absorbed into Microsoft.
- **Runway ML** faces direct pressure from OpenAI's Sora and Google's Veo.

The pattern: when Notion shipped AI writing, every “AI notes” startup lost their reason to exist. When Salesforce shipped Einstein Copilot, AI-for-CRM wrappers lost their pitch. **The platform always wins.**

The deeper insight: **AI is an ingredient, not a product.** Like electricity. You don't sell electricity, you sell the refrigerator. The durable AI businesses are the ones where AI reduces the cost of delivering a workflow that was already valuable. Workday using AI to surface payroll anomalies. ServiceNow using AI to triage tickets. They had the workflow, the integrations, the contracts — AI made the product better, not the business model.

Chatbots solve a UI problem and call it a data problem. The actual value in customer support isn't the chat interface. It's: *what are customers asking, what patterns exist, what does that tell you about product failures?* The data underneath the chat is where the moat lives.

Where the real money is

While AI wrappers commoditize, the data infrastructure layer is consolidating into durable businesses.

Infrastructure (product companies):

Company	Revenue/Valuation	Moat
Snowflake	\$3.6B revenue, 29% YoY	Consumption model, migration switching cost
Databricks	\$1.6B+ ARR, \$43B valuation	Open-source ecosystem (Spark/Delta Lake)
Fivetran	~\$5.6B valuation	Hundreds of maintained connectors — <i>maintenance is the moat</i>
dbt Labs	~\$4.2B valuation	Transformation lingua franca, community lock-in

Services (implementation layer):

Firm	Revenue	Model
Slalom	~\$3.8B	Data & analytics practice, ~15-20% EBITDA
Avanade	~\$3B+	Deep Azure/Fabric integration, certification-gated
Credera	~\$500M (est.)	Boutique strategy-to-execution, higher day rates

The economics of data consulting work: senior data engineers bill at \$175-275/hr in the US. A 10-person team at 75% utilization generates \$5-8M revenue/year. At 25% net margin: \$1.25-2M profit. Not venture-scale, but a genuinely good business.

Boutique plays (5-20 person shops) specializing in a single platform (Snowflake, dbt, Looker) or vertical (healthcare data, retail analytics) regularly hit \$2-5M revenue with 30-40% margins and zero venture funding. They're not

on TechCrunch. They sell — often to PE rollups consolidating the fragmented consulting market.

The pattern across infrastructure and services: **the durable winners own a layer of the stack that's painful to replicate or replace.** Snowflake owns the compute. Fivetran owns the connectors. dbt owns the transformation language. Good consultancies own the client relationship and the institutional knowledge of a specific data environment.

The losers tried to own the interface. The winners own the plumbing.

Where this thesis breaks down

Three honest counter-arguments worth considering:

AI might automate the work itself. Microsoft Fabric's Copilot, Google's Gemini-powered data agents, and dbt Cloud's AI assistant are explicitly targeting the ETL/integration layer — the exact work a data consulting firm would charge for. AI handles clean, well-documented sources well today. It still struggles with legacy ERP exports, inconsistent field naming, and undocumented business logic baked into 15-year-old SQL. The runway is probably 3-5 years, not 10. A business built on manual data integration in 2026 needs a fast pivot plan.

Pure consulting doesn't compound. A 10-person firm billing at \$200/hr is a lifestyle business, not a scalable one. The “boring business” framing works for laundromats or waste management because margins are stable and capex is the moat. Consulting has neither. The productized IP path — proprietary connectors, a vertical data model, a reusable monitoring layer — is the only way out of the hours trap.

Mid-market companies may not pay. A \$100M manufacturing company *should* invest, but often won't. The pain isn't acute enough to displace other priorities, the internal champion lacks political capital, procurement cycles are slow.

Companies that reliably pay are \$500M+ with a CDO who has a mandate. Below \$200M, you're often selling to someone who doesn't have a budget code for what you're offering.

These don't kill the thesis. They sharpen it.

The strongest version of the play

A vertically-focused, productized data consultancy that:

1. Picks one or two industries with high fragmentation AND budget — utilities, healthcare, construction
2. Builds reusable data models and pre-built connectors for the common stack in that vertical
3. Uses an “AI readiness audit” as a low-friction wedge to get in the door
4. Solves the data plumbing problem with consulting plus productized tooling
5. Layers AI applications on top of clean data — but only after the foundation is built
6. Has a 3-year horizon before re-evaluating whether AI automation changes the model

The biggest risk is building a consulting practice that doesn't compound. The mitigation is productized IP from day one — every client engagement should produce reusable assets, not just billable hours. If you're just selling time, you're an agency. If you're building a library of vertical-specific data models and connectors that get better with each engagement, you're building a boring business.

The timing argument

2026 might be the best possible entry point.

AI hype has created urgency: *we need to use AI*. Data reality has created the blocker: *our data isn't ready*. The gap between aspiration and capability is where the money is.

In 3-5 years, AI agents may close some of this gap. But by then, the vertical expertise and client relationships are the moat — not the plumbing work itself.

The winners of the AI era won't be the ones who built the prettiest chatbots. They'll be the ones who quietly fixed the data underneath, embedded themselves into workflows, and were sitting in the room when their customers were finally ready to buy AI that actually worked.

Boring. Durable. Compounding.

The way the best businesses have always been built.

Built by Oliver Signorini

Find more at oliversignorini.com