

The context gap

We read 10 industry reports (and surrendered our email addresses to 10 different mailing lists) so you don't have to. Here's why they all think data readiness is the biggest blocker to AI development.

Let's agree on what 'data readiness' means

42% of teams say more than half of their AI projects have stalled — and the cause isn't the model. It's a lack of data readiness.

— FIVETRAN, AI & DATA READINESS REPORT, 2025

Everyone is talking about how important 'data readiness' is, but no one agrees on what that actually means. So if you're confused on how to get your 'data ready', you're not alone. The challenge is, each vendor is specialized in their piece of the stack and are defining it through their lens: clean tables, pipeline uptime, a governance policy.

In reality, it's all of those things, because they all impact how easily you and your agents access and understand the data. So we've written a new definition of data readiness for everyone to get behind: **All of your data in one place that is reachable by you and your agents.**

Agents fail because they have no idea what's going on

The average company now runs 106 SaaS apps, each one throwing off its own data.

— BETTERCLOUD, SAAS STATISTICS, 2025

*This is about 100 more
than anyone needs*

An agent is only as good as the data it can access. Give it product events but not revenue data, and it can't tell a churn risk from a power user. Give it support tickets but not usage, and it guesses at cause. The problem is that most users don't know when their agent fails. The agent will keep giving you confident, wrong answers until the team stops trusting it.

○ AI-driven acceleration is outpacing trust and governance, and 56% of teams name data quality as their most frequent challenge.

- DBT LABS, STATE OF ANALYTICS ENGINEERING, 2025

This is where AI product development stalls. Teams blame the agent but the real problem is that the agent just had no idea what was really going on in your business.

○ 2/3 data professionals admit they don't completely trust the data their AI is built on.

- MONTE CARLO, STATE OF DATA QUALITY SURVEY, 2023

← The other 1/3 are probably in denial!

Pipelines are the primary problem

AI can't see the full picture so teams try to unify the context themselves. That means pipelines: extract from each source, load it somewhere, model it, then keep it in sync when a schema changes upstream.

○ Over 80% of teams use AI to build pipelines, but overwhelmingly report generic AI tools aren't up to the task, citing hallucinations and missing context.

- ASTRONOMER, STATE OF APACHE AIRFLOW 2026

Pipelines are the tax on data readiness, costing time and money to maintain that could be better spent building something new. Because the data tools on the market are mature but highly specialized, you end up needing 6 different data tools to get the answers you need. And you need pipelines to move data between them all.

○ 70% rate pipeline management as complex, and 64% of data teams spend more than half their time on repetitive tasks.

- MATILLION, DATA INTEGRATION & AI-READINESS REPORT, 2025

How to close the context gap (what every data vendor is trying to tell you)

You need a way to simplify building and maintaining pipelines **so that your data can be in one place** that is reachable by you and your agents. Which coincidentally, is our new definition of data readiness.

Information is trapped in silos and data management is split across many different tools, causing teams to struggle to accelerate AI.

- DATABRICKS, STATE OF DATA + AI

All the reports point teams to a context warehouse which wraps pipelines, modeling, storage, and querying in one tool instead of four you wire together.

hmm...what a coincidence

At PostHog, this looks like product data that lands as it's captured, business context that syncs in from external sources, and those pipelines being maintained automatically. As well as every PostHog tool — analytics, experiments, flags, agents — running on that same data, so context is unified by default instead of assembled by hand.

The gap closes, and your product development kickstarts not because you built a better pipeline, but because you stopped needing one.

Data readiness comes down to having all your context in one place,
ready for you to act on it.

See how the PostHog Context Warehouse works →

posthog.com/context-warehouse

*Can we make this
CTA "Pop" more?*