



As CIOs and other tech leaders face pressure to adopt AI, many organizations are still skipping a crucial first step for successful deployments: putting their data house in order.

Despite warnings going back [at least six years](#), many CIOs fail to collect and organize the vast amount of data their organizations continuously generate, according to some data management vendors. Less than half of organizations have a coherent data management process in place before they launch AI projects, say IT leaders at Databricks and Astera Software, both in the data management space.

Only about 20% of organizations have data strategies mature enough to take full advantage of most AI tools, estimates Naveen Rao, vice president of AI at Databricks, a data management vendor that keeps popping up in successful AI projects. Some small AI projects can work based on a limited amount of company data, or data from outside the company, but many successful AI deployments require comprehensive internal data, he says.

A lot of what we do today when we talk to customers about generative AI is actually level set what's possible," he adds. "If they don't actually have their data in order, they're not going to have the impact they want."

## Pressure to launch

Meanwhile, less than half of organizations have data strategies in place to support any kind of AI deployments, adds Jay Mishra, COO at Astera Software, another data management vendor. Some organizations have little concept of data management, but still are launching AI projects.

"There is a lot of pressure from investors, from the market, to go into AI," he says. "They start with something, and after spending a few months they realize that it has not given the desired results."

If IT infrastructure and computing power make up the engine of AI, data is the fuel, adds Jeff Boudreau, chief AI officer at Dell Technologies. "Even the most sophisticated AI applications rely on quality data to function," he says. "Data is the differentiator. Bad data equals bad AI."

The data maturity observations from Rao and Mishra match, in some ways, a [recent survey from Gartner](#). Sixty-one percent of chief data and analytics officers surveyed agreed that ChatGPT and other technology market disruptions forced them to evolve or rethink their data and analytics strategies.

However, 78% of CDAOs said their data and analytics strategies evolved enough during 2023 to support innovation. It's likely, however, that companies with CDAO or chief data officer roles are ahead of the data management curve.

## Common data problems

Data management challenges come in four buckets:

- First, data exists in silos. The marketing team's data may reside in a different location, with different access rules, than the engineering team's data.
- Secondly, most organizations have generated tons of data, and they're creating more every day. Without a data management plan and system, the old data is buried in folders in a dark corner of an old server, and the new data isn't being cataloged and organized.
- Data is incomplete, inaccurate, and inconsistent.
- Finally, a large percentage of data is unstructured and therefore, isn't easy to organize. Crucial data resides in hundreds of emails sent and received every day, on spreadsheets, in PowerPoint presentations, on videos, in pictures, in reports with graphs, in text documents, on web pages, in purchase orders, in utility bills, and on PDFs.

Text documents, often stored in multiple locations across an organization, often contain a wealth of information, Astera's Mishra says. An important data point could be buried on a chart on page 5 of a 20-page document, or throughout a 100-page Wall Street analyst report.

"A lot of data that is produced by the regular application or business users stays in documents, and documents are still the biggest form of communication," he says. "That data is free flowing and does not reside in one place. That's a huge challenge and opportunity."

## More data doesn't always produce better AI

One misconception about the volume of data that companies hold is that feeding AI models more data creates better AI results, Mishra adds. While some AI tools demand high volumes of data, quality is more important.

"The data that is not curated is going to be the basis for the wrong results," he says. "The quality of data determines everything."

But AI users shouldn't discount the demand for data from large-language model AIs, says Bryan Eckle, CTO at cBEYONData, a professional services provider for US government agencies.

“AI is very, very data hungry,” says Eckle, who evaluates AI tools for customers. “And the data needs to be accurate, it needs to be timely, it needs to be fast, and there needs to be a lot of it.”

Beyond the four big buckets of data management problems, organizations also struggle with a single source of truth in their data, Eckle says. Which of the five versions of a product specification PDF floating around an organization is the correct one? Does your customer support chatbot have access to all five versions?

## **Focus on quality and standardization**

For those organizations struggling to clean up their data, Dell's Boudreau recommends focusing on data management processes and [governance](#) that consider privacy, standardization, quality, and integration.

Even before organizations start to clean up and organize their data, Eckle recommends they think through their goals for the data.

“You can back up and start with, ‘What kind of questions do we want to be able to answer?’” he says. “Then from there, ‘What are the underlying data elements we need in place in order to answer those questions?’ And then from there, ‘What's the source of truth?’”

Cleaning up the data is often ignored in AI projects because it isn't the flashy piece, Eckle adds. But a huge part of an AI project, 80% or more, is cleaning the data.