

How Data Drives Generative AI Success

Leadership strategies to build greater data quality and maturity and foster a data-driven culture

Table of contents

Introduction 3

The Cost of Technical Debt from Legacy Data 4

Three Phases of Data Maturity 5

- Data-informed
- Data-aware
- Data-driven

Building a Data-Driven Culture to Support Generative AI 7

- Step 1: Find the “data huggers”
- Step 2: Debunk the “good enough” myth
- Step 3: Challenge statements based purely on intuition

Generative AI–Ready Data: A Checklist 11

Keeping Data Foundations Strong as Generative AI Usage Grows 12

Conclusion 13

Introduction

Realizing the full potential of generative AI starts when everyone learns to channel their inner chief data officer (CDO).

Even if you have an actual CDO in your organization, managing and using data strategically should never fall to just one person. It's everyone's job, because data is what fuels technology like generative AI to transform everyday work. For example, developers who want to build generative AI virtual assistants need accurate data about products, services, and company policies. Marketing teams need quality data on customer needs in order to generate content.

It takes leadership and a collective effort to establish solid data foundations. And now is the time to get the fundamentals right. Organizations must identify their data sources, how it's consumed, and by whom. The data needs to be high quality and processes are needed to measure and enrich the data quality. This means treating and managing data as a product, because it's only useful if it's accessible and used by those who need it.

To democratize data, it must be discoverable, ideally through self-serve capabilities. And to protect data, it needs to be classified by sensitivity and risk, and organizations must have strong controls and policies in place.

The first step is addressing the current state of data and recognizing why it often won't support what organizations want to do with generative AI.

How are organizations currently using their data?

WORLDWIDE ORGANIZATIONS

40% of organizations worldwide say they increased value from data to a high degree. **(source)**

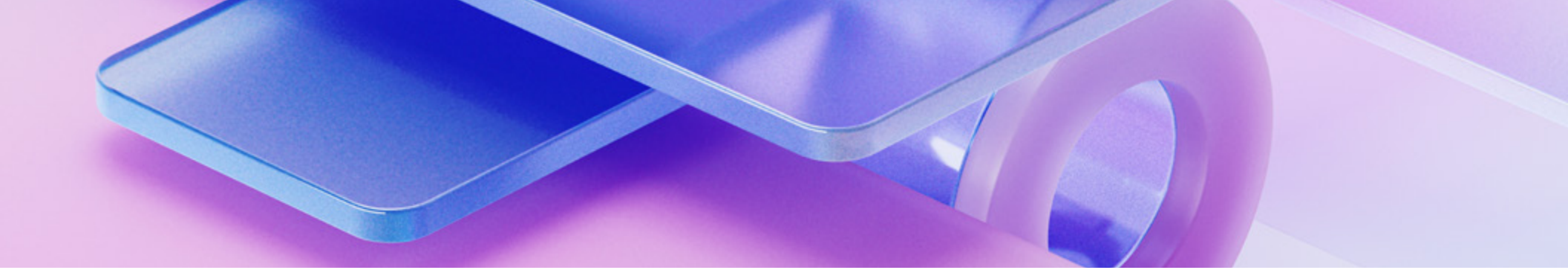
CANADIAN ORGANIZATIONS

22% use data for strategic use cases like building customer profiles and personalizing service. **(source)**

15% use proprietary data in generative AI tools. **(source)**

13% use customer information such as their names.

49% always check whether generative AI outputs from this data are accurate.



The Cost of Technical Debt from Legacy Data

It was easy to see how quickly data accumulated when it was stored on paper. What started in a filing cabinet would move to archiving boxes and fill up an office's storage facilities. Though digitization made data far more accessible and usable, it didn't automatically address the problems that ongoing accumulation creates.

Sales data is now stored in CRM systems. Marketing data in customer data platforms. HR keeps its own data, as do finance, legal, and operations departments. **The average organization now uses 976 applications** to manage all this data. That means it is often siloed, making it difficult to combine and get answers to complex, mission-critical questions.

Legacy data is not just fragmented by being buried in old software. It is also often:

- Unstructured and unorganized
- Stored in outdated formats
- Incomplete, inaccurate, or irrelevant

This is an often-overlooked form of technical debt. It's not unlike when companies struggle to modernize their IT because they first have to deal with aging hardware. Old servers and network equipment can be a barrier to migrating onto the cloud and using advanced applications like generative AI.

Addressing the technical debt from legacy data doesn't happen overnight. It requires leaders to learn the right path that will move organizations forward in terms of mature data practices.



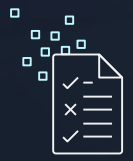
DATA MATURITY

Three Phases of Data Maturity

Generative AI lets organizations create new value from data, but a lot depends on how they collect, store, and manage it.

1 IN 3 Canadian organizations says its data management and data governance processess are in the mid- to high-maturity range. **(source)**

Reflect on how far your organization has progressed based on these three stages:



Data-informed



Data-aware



Data-driven



Data-informed

An organization that focuses largely on historical or lagging data, such as sales figures. The emphasis is on backward-looking data that senior executives can use to develop the organization's strategic direction.

These "data-informed" organizations are limited by an inability to use data other than to estimate future events and base planning on previous results. This type of data must also accumulate over time—weeks, months, or even years—in order to have a statistically significant amount of information to identify trends for decision-making.

Data-aware

As organizations develop dashboards and data warehouses, they use applications like business intelligence software to monitor performance and optimize processes. This typically involves putting data into batches and relying on IT departments to perform queries that give senior decision-makers the insights they need.

These "data-aware" organizations can often struggle because they are still working in silos and may not be getting the full picture from other areas of the business. They're also at the mercy of the IT department's availability to make updates or changes to the reporting.

Data-driven

Organizations that combine clean, well-structured, and real-time data with technologies like AI can guide and even trigger actions. A "data-driven" organization democratizes data. That means making it available to employees in accordance with data access policies, and prioritizing it in shaping customer experiences.

Becoming a data-driven organization isn't just a technology issue. It's a mandate to implement a data operating model and leading employees to embrace a data culture.



Building a Data-Driven Culture to Support Generative AI

A full 66% of Canadian organizations say using generative AI will help them take better advantage of available data. That number will be even higher once more leaders guide their team toward a data culture.

In a data culture, employees learn to think differently about data. They become more comfortable working with it and value the different ways it can be used. To get there:

1. **Find the data "huggers"**
2. **Debunk the "good enough" myth**
3. **Challenge statements based purely on intuition**

1 Find the “data huggers”

Data doesn’t get fragmented or siloed just because of all the platforms and applications that organizations adopt. Part of it stems from “data hugging,” where people working in a specific department have control over a set of data and resist sharing that control with others.

Data huggers usually have a good reason for their behaviour. They’re concerned about privacy and security risks if data becomes more accessible. They may already know that the data lacks sufficient quality, but only they understand what the quality issues are. Finally, they’re worried that quality issues will get worse if the data is shared across departments and used in tools like generative AI.

All of these objections can and should be addressed. Here’s how:



Implement privacy by design and ensure proper data protection measures are in place. Employees may not be aware of the tools and processes that are available to mitigate cybersecurity and privacy risks. Educate and train them on how to share data safely and without compromising privacy, with clear controls over who has authorization and access to use it in generative AI applications.



Help employees see themselves as data stewards, rather than gatekeepers. Their role is to make data available, and their performance should be evaluated based on their ability to help others derive value from data through tools like generative AI. Align their incentives and rewards to support this behaviour, recognizing them as they shift toward a data-driven mindset.



Involve employees in the process for adopting generative AI. Data hugging will look a lot less attractive once team members understand how generative AI lets them work faster and achieve better results. Expose them to the use cases that are being developed so they buy into the process and appreciate the role that data plays.

2 Debunk the ”good enough” myth

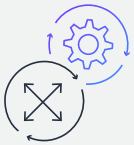
Poor-quality data becomes acceptable in organizations when employees see it solely as a means of facilitating transactions, such as tracking the number of products sold.

As long as the data—or even parts of it—is clean enough to be able to accomplish what an employee or team wants to do with it, the deeper quality issues are ignored. The data might record sales transactions, for instance, but doesn’t sync properly with the customer’s profile or account history.

Once the data moves downstream to be used in generative AI, “good enough” often leads to poor outputs. Don’t let that happen:



Provide the right tools and processes and teach employees to treat data as a corporate asset. If the company owns an office building, point out that failing to clean and maintain it regularly would reduce its value and make it more difficult for employees to do their jobs. The same is true for corporate assets like intellectual property: if trademarks and copyrights aren’t renewed, the company risks losing its key differentiators. Data is no different. If data quality is poor, the asset’s value drops, which affects everyone’s ability to accomplish their goals.



Make sure data is well utilized. It’s easy to overlook quality problems when data is sitting untouched in repositories for prolonged periods. Set a schedule to conduct regular reviews of data assets to make sure that it’s actually being used and still needs to be stored.



3 Challenge statements based purely on intuition

A data-driven organization needs to be action-oriented, but its people should base actions on concrete metrics. Coach employees to transcend decisions based on anecdotal information or their feelings and instead anchor them in data.

This isn't to discredit longstanding employees with a lot of institutional memory and experience. Encouraging the rigorous use of data should be seen as a way of helping them build upon what they've learned first-hand.

Here's a few examples of the kind of statements that come up in everyday meetings and in messages shared across the organization:



"We've substantially increased our revenues."
That may be true, but ask for a precise percentage based on data from the CRM program or other source.



"Customers are much happier with this way of doing things."
There are always opportunities to improve the experience that you're delivering, but press employees to stick to data-based metrics. What has the change in process meant in terms of customer satisfaction (CSAT) or net promoter score (NPS)?

Promoting data literacy enriches everyone's vocabulary as teams collaborate to solve problems and make decisions. It encourages everyone across the entire organization to treat data as a valuable and valued commodity. This also prepares employees to develop good habits, like carefully checking the results of what they produce through generative AI.

Generative AI–Ready Data: A Checklist

These are the characteristics that will show you're moving in the right direction:

Data is harmonized

Whether data originated in an ERP system, a customer service application, or a survey, data harmonization means that employees are able to connect the dots and find insights. The data has been cleaned, has been mapped to a common schema, and remains accurate no matter where or when someone accesses it.

A semantic model has been applied

Bringing data together is a great first step, but it's not enough. A semantic model means that you've created a layer on top of the data that promotes a shared understanding of what it means. This includes the relationship between different datasets.

Data insights have been democratized

The ability to make sense of data and act on it should not be limited to data engineers or data scientists. Data should be available to those who need it. Anyone in the organization should be able to query the data using low-code or no-code tools, or even natural language queries through generative AI.

Data doesn't always have to move

Organizations have traditionally used a complex, costly process called extract, transform, load (ETL) to put data in a central repository. Talk to the IT team about using a zero-ETL framework, which allows employees to query data from its original source for greater agility, efficiency, and real-time insights.

Employees are confident in the data that they're using for generative AI

Add questions to existing employee engagement or pulse surveys to gauge how the team perceives data quality, accessibility, and governance. Get more feedback through employee town halls or even team huddles between managers and staff.

Governance and security processes are continually improved

Concerns around bias and ethical use of generative AI will require ongoing work to put the proper guardrails in place. This should be more than an agenda item for the CISO. Organizations should regularly update training and education for all employees around the risks and best practices in responsible data use.

Keeping Data Foundations Strong as Generative AI Usage Grows

Getting data ready for generative AI should be the beginning of an ongoing organizational effort. Empower employees to build upon the data culture that's been established by continually improving key processes and policies.

Gather diverse datasets and augment where necessary: As organizations discover and pursue more generative AI use cases, they will need to collect new forms or additional sources of data. This can range from news articles and technical documents to images, audio or video files, or a combination of formats. Make identifying these needs one of the first steps in generative AI pilot projects.

Document data lineage: Data lineage tracks where data came from, how it has been transformed, its relationship to other datasets, and where it flows. It establishes that what's being used is complete, accurate, and trustworthy. The metadata stemming from data lineage also supports data discovery, data quality, and data access control implementation.

Version data used in models: Generative AI can and should lead to experimentation with foundation models. Along the way, that data will obviously change. Where AI models are customized with an organization's own data, ensure the ML Ops processes have traceability to track the model version and the data used for training the model.

Aim for speed via automation: Capitalizing on generative AI is already becoming competitive among organizations in every sector. Reduce or eliminate the manual work involved in getting data ready. Use tools that can collect, process, transform, and analyze data far more quickly than a team of humans. Generative AI tools can accelerate with the analysis and automation efforts.

Be aware of and manage costs: Organizations don't have to shoulder the expense of developing generative AI on their own. Take advantage of foundation models that can be customized with your own data, or apply techniques like Retrieval Augmented Generation to create differentiated experiences.



Conclusion

Organizations see the potential for generative AI, and with good reason. It's a technology that offers almost limitless possibilities in driving greater productivity, higher efficiency, lower costs, and enhanced customer experiences.

A thoughtful, comprehensive data strategy will accelerate the time to value for generative AI by avoiding bottlenecks and costly errors.

More importantly, it will help employees to develop the skills that they need to maximize the technology's potential. And as they use generative AI, they will have greater confidence in its ability to serve customers better in the moments that matter.

About AWS

With our deep AI expertise and over 100,000 customers, only AWS provides the most comprehensive set of services, tools, and resources to meet your business needs. From builders to buyers; from data scientists to business analysts; from students to AI experts – AWS makes AI accessible to more people, enabling the next generation to build innovative, new technology.

AWS offers the easiest way to build and scale generative AI applications, grounded in model choice and flexibility, making it possible for builders of all skill levels and organizations of all sizes to transform their business. Built on over 20 years of AI experience at Amazon, AWS has helped enterprises, start-ups, and non-profits unlock the value of all their data with AI across every line of business, with end-to-end security, privacy, and AI governance.

Learn more

