



DATA

What's Your Data Strategy in the AI World?



Daniel Viveiros
SVP of Data & Analytics, CI&T

Introduction

2023 has been a busy year in the data space. The excitement surrounding Generative AI and its protagonist, ChatGPT, has shed light on something not entirely new but largely dormant or underutilized by many companies: Artificial Intelligence, in general. So, why has this topic suddenly surged in popularity once again? Let's explore a couple of compelling reasons.

First, AI has evolved beyond merely finding patterns, bringing insights, and predicting future trends. It now has the ability to create new content (text, video, and image) based on accumulated knowledge. And the results are pretty impressive. Second, OpenAI has made it extremely easy for people to get value from their products. With a straightforward chat interface, albeit poorly named ChatGPT, one can significantly boost personal productivity. Their APIs are also seamlessly integrated with other systems, creating an AI platform with intrinsic value, ideal for accelerating business efficiency and enhancing customer experiences.

Is that it? Have we figured out the extent of AI's potential with ChatGPT and Generative AI? Roy Amara, an American scientist and futurist, coined Amara's law, which cautions us: "We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run." Many people think that Gen AI will change every aspect of our personal lives and business, disrupting everything we know. That's probably the overestimating part of the law. While it will undoubtedly increase productivity and disrupt certain industries, particularly those heavily reliant on content creation, the long-term winners will be those with a comprehensive and connected data strategy. They will blend pre-trained models with their own datasets, leverage different AI techniques, build robust platforms to store/serve data at scale, and - above all things - connect everything with revenue streams and cost reduction opportunities.



Biggest Challenges to Data & Analytics Programs

It takes work to define and deploy a data strategy. According to this article on MIT Management, only half of the CDOs (Chief Data Officers) are able to drive innovation using data, and roughly 26% of them have succeeded in creating a data-driven organization. In our experience, the most significant pitfalls for a data initiative are:

1. Data not solving a business problem

We can't stress this enough. This is the most common anti-pattern for a data program. Companies often prioritize technology and engineering, delaying the critical discussion of how to use data for business impact. The belief is that creating a data lake in the cloud with tons of data sets is the journey's first step. Practice shows that modern architectures are instrumental to success, but knowing the problem to be solved and defining an operational model connected to business areas should come first. Our recommendation is to design and validate the target architecture. The implementation should be pulled by business needs. The same thinking is to governance. The value streams prioritize specific tasks and details after establishing the first set of guidelines and guardrails.

2. Introducing complexity too soon

Crawl, walk, run. This approach has been used many times in the technology industry. We still see many companies introducing complexity too soon in the process. It can be in the data architecture realm or the data science one. Complex and multilayered data lake architectures are being implemented before actual data exists to be handled and served. It's a good idea to design the future state of your data architecture. The execution can be incremental, enabling business capabilities as quickly as possible. The most essential step for AI/ML models is to have a baseline for relevant metrics. It can be simple and straightforward to serve as the first comparison basis. Future iterations will compare results against the baseline and redefine it. Complexity will be added incrementally alongside the results achieved by each step.

Biggest Challenges to Data & Analytics Programs

3. Data in silos, with quality issues

Creating and maintaining high-quality data repositories takes a lot of work, especially in large and complex enterprises. Departments and product lines have their own data sets. M&A activities create significant complexities and overlaps in terms of systems of records. Data is inaccurate, incomplete, outdated, duplicated, or inconsistent in many scenarios. No AI strategy will take place or survive without high-quality data sets.

4. Neglecting UX and people factors in data programs

Analysis and insights are great, but the actual value for your business will come from actions. Most of the time, you'll rely on people and good UX approaches to unlock outcomes. Let's take an order recommendation algorithm, for instance. How are you going to approach your online user with the recommendation? In which part of the journey does the user have a higher propensity to accept the output of the algorithm? And the challenge goes beyond digital experiences. In a B2B scenario, the order recommendation may target a salesperson, usually compensated with sales commissions. Will they trust the recommendation? This last scenario is a perfect example of how critical is to consider the user journey in your AI strategy. Your company can have a bulletproof AI model, which will be useless if people don't feel safe consuming the outputs.

5. Balancing agility, democratization, and governance is essential

While promoting autonomy in business areas enhances agility and democratization, it must be balanced with corporate strategy and governance to ensure compliance, security, and data privacy regulations. Quality data and well-designed architectures are crucial for AI strategies, necessitating effective data governance that goes beyond control to foster a data-driven culture.

Four Data Pillars for a Solid Analytics & AI strategy

Here are the four pillars to consider when building and implementing your data strategy:



STRATEGY & GOVERNANCE

- What's our data strategy (in an AI world)?
- What are the business imperatives for data?
- How can we manage data practices across different business units? How do we enforce security and privacy guardrails?



ENABLEMENT & INNOVATION

- What's our architecture to handle data at scale with low latency?
- How to create and deploy efficient data pipelines?
- How to improve data quality and automation over time?



LITERACY & DEMOCRATIZATION

- How to promote data-driven approaches to non-data people?
- How to empower product teams to use data?
- How can we help our people to understand, interpret, analyze, and communicate data effectively?



PLATFORM & OPERATIONS

- How can we improve revenue connecting data to our decisions and customer experiences?
- How can we maximize efficiency leveraging data?
- How can we spot and capture new opportunities with data products?

All of these aspects are critical for a successful data program. We've observed initiatives that lacked one or more of these components, leading to various short-term and long-term issues. Here are the most common pitfalls:

Excessive focus on engineering: Some companies invest heavily in the Platform & Operations pillar, postponing business discussions to later phases. The belief is that scalable data architectures and vast amounts of data are the first step towards a data-driven culture. Success is measured by data availability and scalability. The blind spot here is that success is actually measured by business outcomes. And nothing promotes change better than success. Even if your company needs to modernize its data foundation, it should be a yellow flag taking more than 1-3 months to bring business to the table.

Excessive power-to-the-edge. On the other side, some organizations run disconnected data initiatives across various business areas, emphasizing the Enablement & Innovation pillar. This approach may yield quick initial results, but it often leads to data silos, weak commercial agreements with data platform providers, and compliance challenges with regulations like CCPA/GDPR. Over time, data quality may deteriorate, and technical debt can accumulate.

Governance as a control mechanism... only. Governance can be a powerful tool to promote practices and, ultimately, help create a data-driven culture. But this happens only when the agenda for governance goes beyond control and is an enabler for business outcomes. In other words, the Strategy & Governance pillar must closely connect with Enablement & Innovation, guiding the Platform & Operations backlog. In a standalone mode, this pillar works as a handbrake for innovation and cultural change.

Lack of data-driven culture. Data matters to data people, right? Wrong. Data has absolutely no value if not connected to business outcomes. This connection can't rely solely on data teams; it must involve product teams, business professionals, and product managers. These individuals are experts in their respective fields and can identify opportunities to use data to enhance performance. Basic data literacy, including knowledge of ML/AI concepts, is crucial for fostering a data-driven culture within the company. This is where Literacy & Democratization play a pivotal role.

At this point, you've realized that you'll need all of them. The good news is that you don't have to master every capability from each pillar. The data journey is an incremental process. New capabilities are pulled by business needs, with governance and cross-pollination strategies to replicate practices in other areas. In our experience, the most successful pattern is to develop the data program in the following order:

Enablement & Innovation

- Select a few (a maximum of three) business areas that could benefit from Data & Analytics (including AI/ML).
- Map the value stream, identifying friction points, bottlenecks, and opportunities.
- Create hypotheses on how data can help.
- Define success by specifying the business metrics to be measured.

Literacy & Democratization

- Given the prioritized areas, level set the knowledge on data sets and practices.
- Immerse data people (scientists, engineers) into the product teams.
- Co-create hypotheses and solutions.
- Establish a unified vocabulary to measure and communicate results based on the selected business metrics.

Strategy & Governance

- Identify leverage points in the roadmap with existing solutions.
- Define necessary guardrails.
- Incorporate learnings and data into the knowledge base.
- Promote approaches and results to executives and the community.

Platform & Operations

- Evolve the data architecture to enable the business initiatives.
- Incorporate data sets and create/review data pipelines.
- Ensure data quality and observability; automate as many data processes as possible.

Of course, this is not a waterfall project. You don't have to "complete" one pillar to move to the next one. Conversations must happen in parallel. The proposed order is to emphasize the drivers and mindset of the program: from business to technology.

Capabilities and Practices for Each Pillar

Each company must define its own roadmap for data practices and capabilities. In our experience, we have the most important ones for each pillar. Again, the list below isn't supposed to be exhaustive but to highlight concepts we see more commonly requested and deployed on data initiatives.



STRATEGY & GOVERNANCE

Data/AI Strategy

Data Ownership

Data as a Product

Master Data Mgmt

Data Governance



ENABLEMENT & INNOVATION

Business Intelligence

Advanced Analytics

Customer Analytics

Data Monetization

Gen AI: Efficiency & CX



LITERACY & DEMOCRATIZATION

Data Literacy

Community of Practice

Data Training

Self-service BI/Platform

Data Privacy & Security Education



PLATFORM & OPERATIONS

Data Architecture

Data Quality

Data Pipelines / ETL

DataOps/MLOps

Data Observability

Conclusion

AI has a long history of transforming industries, and this trend will continue, but at an accelerated pace. AI/ML APIs, auto-ML tools, and pre-trained models are democratizing AI, presenting two significant challenges for the future. Firstly, companies must create differentiation by leveraging their own data to craft unique solutions. Secondly, they must bridge the gap between technology and business to ensure data and AI enhance performance rather than just serving as marketing tools.

Generative AI generates short-term excitement but, as Amara's Law signifies, long-term success depends on the Data/AI strategy your company defines now. A strategy that should rely on technology, governance, democratization, and, above all, business impact.

CI&T is a digital specialist with a long track record of deploying Data Strategies for companies like Nestlé, AB-InBev, and Bank Itaú. We believe that unlocking your data is the key to making better business decisions. CI&T's experience in strategy and engineering can help your organization glean insights to drive greater impact.



Thank you

**For more information, or to discuss the content of this paper,
please get in touch with:**

Daniel Viveiros – SVP of Data & Analytics



[Get in touch – LinkedIn](#)

Data at CI&T