

THE QUEEN OVER THE PAWNS

Why Data Strategy Is About Choosing What to Ignore

A strategic lens on AI-era data investment — informed by Porter, chess, and the cost of doing everything

Executive Summary

For more than a decade, enterprises have been told that data is the new oil — and they believed it. Data lakes were built. IoT streams were captured. Every table, every log file, every inventory movement was ingested, stored, and tagged for eventual transformation. The promise was that with enough data, AI would find the gold.

It did not.

What most organizations built was not a strategic asset. It was a very expensive swamp. And today, as AI investments accelerate and CFOs demand returns, the question is no longer how to collect more data — it is how to stop wasting money treating the wrong data as if it were precious.

The most important decision an enterprise can make about data is the same decision Porter argued was central to strategy: choosing what not to do. Applied to data, this means identifying the small subset that directly enables new revenue, new customers, or a defensible competitive advantage — and investing there with full intensity, while deliberately deprioritizing everything else.

The chess analogy is instructive. No grandmaster opens a game by systematically clearing every pawn before advancing major pieces. The player who does that loses — not because they made wrong moves, but because they made too many irrelevant ones. Speed and focus on the pieces that matter is what wins.

The Collectors trap

The decade-long push to collect everything was born from a legitimate insight: you cannot analyze data you do not have. But this insight was applied without a strategic filter, resulting in organizations that now sit on petabytes of data with minimal ROI and rising maintenance costs.

For example, a mid-size manufacturing firm. Among other data they collect: three years of high-frequency vibration sensor data from standard conveyor equipment, daily inventory movement logs across twelve warehouses, raw machine utilization streams from commodity production lines, customer purchase history, CRM interaction logs, and pricing exception data. Often, following the human-ancestor pattern, collecting everything, believing that it is somewhat useful. Just waiting for the right Use case to come up.

Now, under the ever-repeated warning. Garbage-in, garbage out, the scarce data engineering team spending 70% of its capacity cleaning and contextualizing sensor data that no model will ever meaningfully differentiate — because every competitor's conveyor belt vibrates in the same pattern.

"Data is only strategic when it encodes something your competitors cannot easily replicate. Everything else is infrastructure."

The fallacy is not in collecting data. It is in treating uniformity as a virtue — believing that bringing all data to a consistent, clean, contextualized state was itself a strategic activity. It is not. It is janitorial work, and like all janitorial work, it should be done efficiently and cheaply, not extensively and expensively.

Porter's Lens Applied to the Data Estate

Michael Porter's foundational insight was that competitive advantage does not come from doing everything well. It comes from making deliberate trade-offs — choosing an activity set that is coherent, differentiated, and hard to replicate. The essence of strategy, he wrote, is choosing what not to do.

Applied to enterprise data, this translates directly. Ask not "which data should be cleaned?" but "which data, if properly activated, enables outcomes our competitors cannot easily match?" That question produces a very different investment thesis.

THE STRATEGIC DATA TEST

Before committing data engineering resources to any dataset, apply three questions: (1) Does activating this data create a revenue opportunity or protect margin in a way that is difficult to replicate? (2) Does this data encode customer behavior, preference, or relationship that is proprietary to us? (3) Would a competitor with the same dataset achieve the same outcome? If the answer to (3) is yes, you are looking at commodity data — manage it cheaply and move on.

For example, inventory movement data for a standard logistics operation wouldn't pass any of the three tests. It is high-volume, expensive to maintain, and brings insights that any operator with a similar scale can easily gather. Cleaning it obsessively is Porter's trap: mistaking operational efficiency for strategic differentiation.

By contrast, the behavioral signals buried in your customer interaction data — pricing sensitivity, churn precursors, upsell triggers — may pass all three. That data is worth the engineering investment. That is where the AI investment should concentrate.

The Chess Principle: Play the Board, Not the Inventory

Again, the apology Chess helps to illustrate. The beginner might think: "Before I can bring my queen to bear, I should clear the path — remove the pawns, neutralize the minor pieces, establish

control systematically." This sounds very disciplined and straightforward. It is, in practice, how you lose.

Grandmaster's play is not sequential clearing of categories. It is scanning the board for the combination of moves that matter — where this combination creates a definitive, irreversible advantage — and investing tempo there. Pawns are not cleared; they are advanced or sacrificed in service of a larger pattern. The queen is not saved for later; she exerts pressure from the moment the position allows.

"Nobody plays chess by first removing all the opponent's pawns before moving on to the knights. You play to the position — and the position demands that you focus on the pieces that win."

Enterprise data management is often equivalent to pawn-clearing, and this is exactly what most organizations do: normalizing every table before building any model, cleaning every data source before proving any value, and establishing a perfect, unified data layer before activating anything. By the time the foundation is ready, the competitive window has closed.

The right move is to identify the datasets where AI produces disproportionate, defensible value — and to invest engineering capacity there first, in full. The rest of the board can wait.

THE TWO-MOVE PRINCIPLE

A principle often cited in competitive strategy holds that a player allowed to make two moves for every one of their opponent's will defeat any grandmaster — regardless of skill differential. The asymmetry of tempo is that powerful. In data terms, the organization that activates its highest-value data first does not need to have the biggest data lake or the most comprehensive data architecture. It needs to move on the right data while competitors are still debating governance frameworks for the wrong data.

Speed on the right data is worth more than completeness on all data. This is not an argument against data quality — it is an argument against the misallocation of investment in data quality.

What This Means for AI Budget Allocation

The practical implication for organizations planning AI-enabling data investments is a reallocation, not a reduction. The argument here is not that data management budgets are too large — in many cases, they are not large enough in the right places. It is that the distribution of that budget is systematically wrong.

- Redirect data engineering capacity from commodity data pipelines to revenue-generating data activation. The goal is working AI on strategic data within quarters, not years.

- Establish a clear two-tier data management standard: strategic data gets full quality investment; commodity data gets automated pipelines with minimal human curation.
- Retire the myth of the complete data foundation. No enterprise has ever finished building it, and none that waited for it to be ready has won on AI. Start with the queen.
- Measure ROI at the dataset level. If a specific data asset does not contribute to a revenue or retention metric within a defined horizon, downgrade its investment priority.
- Protect tempo. Competitive advantage in AI accrues to the organization that activates strategic data first and iterates fastest — not the one that eventually achieves the most uniform data architecture.

This is not a technology problem. It is a strategy problem. The organizations that will win on AI are not the ones with the most data. They are the ones who identified which data was worth winning on — and moved there decisively while others were still clearing pawns.

The Closing Move

The most powerful data strategy in the AI era is not a data strategy at all — it is a competitive strategy expressed through data. Know which data is your queen. Move it early. Let your competitors spend the next two years aligning their pawns.

"You can beat any chess grandmaster by making 2 moves whenever they make 1 move."