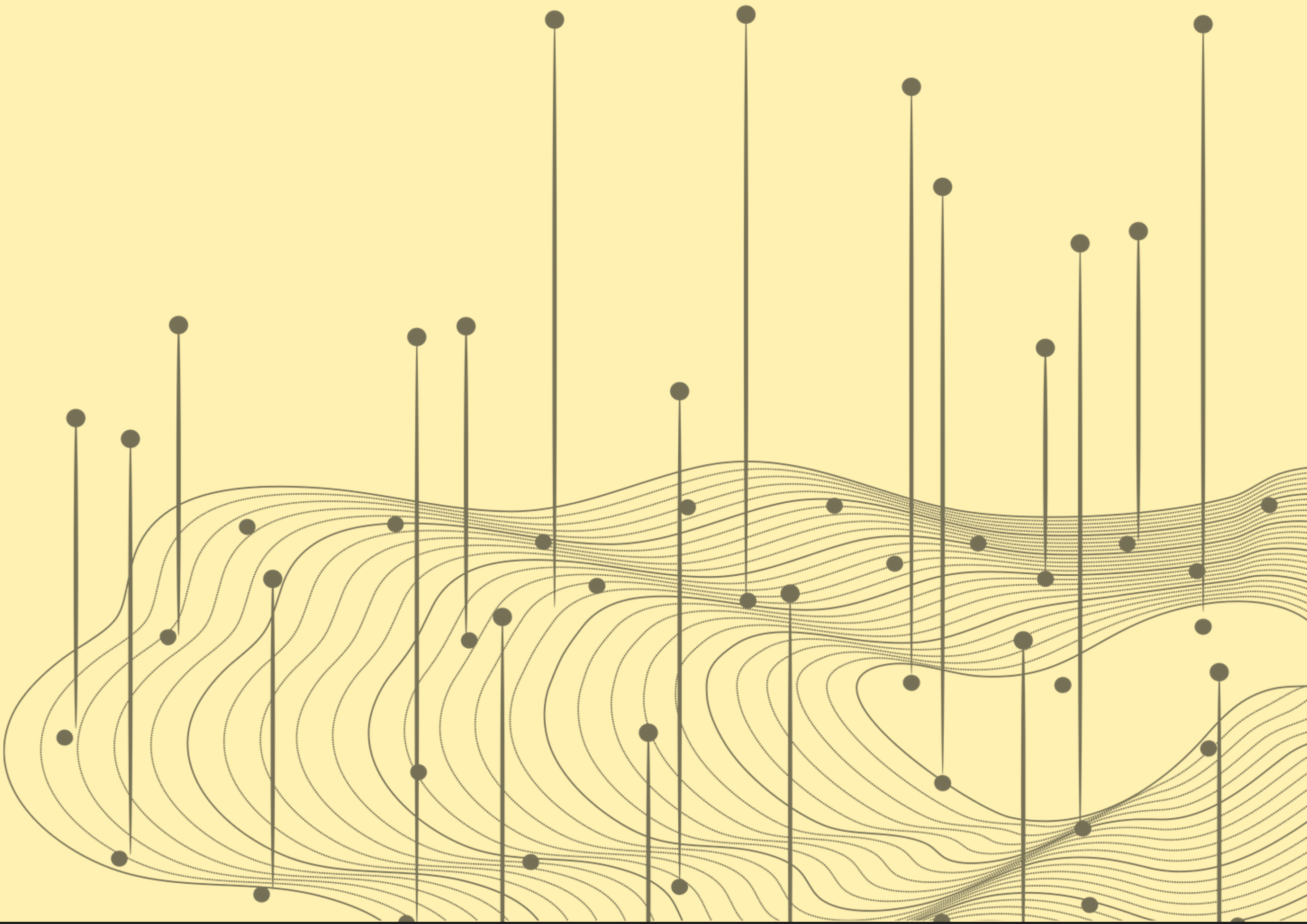


entopy.com

Crossing the data chasm:

Building intelligence before your data is perfect.

entopy

Executive Summary

Artificial Intelligence (AI) has exposed a challenge facing almost every critical infrastructure organisation: data.

Many organisations believe they must first build a perfectly standardised data foundation with a strong, consistent taxonomy before AI and operational intelligence can be deployed. This belief has fuelled significant investment in enterprise data platforms, governance and large-scale transformation programmes, often delaying operational AI by months or even years.

This paper argues that the sequence is wrong. Rather than waiting for perfect data, organisations should begin by solving operational problems. Each successful deployment creates measurable business value, revealing which data actually matters, exposing gaps in quality and creating the organisational momentum needed to improve the wider data estate.

The result is a fundamentally different approach to AI adoption. Instead of treating intelligence as the outcome of data transformation, intelligence becomes the catalyst for it.

We explore four engineering principles that make this possible:

- Dynamic operational ontology: creating a common operational language across fragmented systems without requiring a single enterprise data model.
- Micromodels: breaking complex operations into small, high-resolution prediction problems that require only the data needed for each capability.
- Iterative deployment: delivering intelligence capability by capability, allowing business value to drive future data investment.
- Synthetic data: overcoming scarce and infrequent operational datasets to accelerate model development and improve prediction accuracy.

The central argument is simple:

The fastest route to better data may not be waiting for a perfect data foundation. It may be building better intelligence first.

The AI data dilemma.

AI has become a strategic priority for many leaders across critical infrastructure. But data readiness has become one of, if not the biggest barrier. We don't have enough data. Our systems are legacy and fragmented. We still use spreadsheets. Our data sits in silos. Our systems are not designed to talk to each other. All familiar concerns and, all legitimate. In fact, these concerns are reflected across industry.

But in most cases, the 'data chasm' is often interpreted too broadly. Most organisations do not have a complete absence of data however, many believe for AI to work, you need perfect data: every operational system continuously generates data into a common platform, the data is complete, consistently structured, trusted, secure and immediately ready for analytics and AI.

The dilemma is therefore not whether 'perfect data' is desirable. It undoubtedly is. The dilemma is whether organisations should wait until their data is considered 'AI-ready' before they begin creating intelligence, or whether intelligence itself can become the catalyst for improving data.

The data foundation gold rush.

Faced with this dilemma, the market has largely chosen one path: fix the data first.

If data is perceived to be the barrier, then the logical response is to remove the barrier. Organisations have invested heavily in data platforms, governance, integration and enterprise data foundations. Recognising the opportunity, virtually every major consultancy now offers AI readiness, data modernisation and enterprise transformation programmes, while Enterprise Data Management has grown into a market worth more than US$120 billion.

This approach is understandable. It follows the same model organisations have used for decades to modernise enterprise technology: build the underlying infrastructure first, then deploy new capabilities on top. It also suits longer-term transformation programmes which fits the traditional model.

The unintended consequence is that intelligence becomes the destination rather than the starting point. Operational intelligence is pushed to the end of a large-scale transformation programme, meaning organisations often spend years modernising their data estate before delivering their first meaningful AI capability. By the time they arrive, others have already accumulated years of operational learning, organisational adoption and continuously improving intelligence.

The market has optimised for delivering transformation programmes. Organisations are trying to buy operational intelligence. Those are not always the same thing.

Why perfect becomes the enemy of progress.

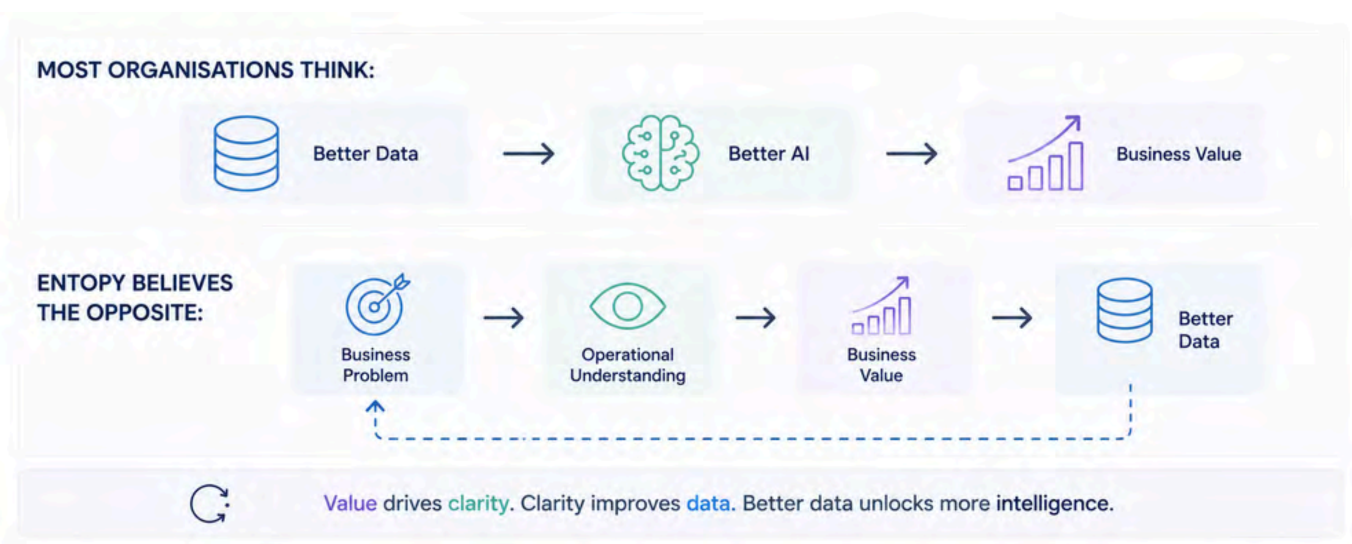
The prevailing view is that better data must come before better intelligence. But we argue the relationship is far more iterative.

Not because data isn't important, but because organisations rarely know which data is truly valuable until it begins improving operational outcomes. Waiting for a perfect data foundation assumes organisations already know which information matters. In reality, they don't. They know they have data. They know it isn't perfect. What they don't know is which datasets will ultimately improve decisions, which relationships matter most and where investment will deliver the greatest operational return.

Rather than attempting to perfect an organisation's entire data estate, this philosophy begins by solving operational problems. Each successful deployment creates measurable business value. That value reveals which data actually matters, exposes gaps in quality, identifies missing relationships and directs future investment towards the information that demonstrably improves operational outcomes.

It also creates something equally important: organisational momentum. Operational teams become advocates because they can see the value being delivered. Data owners become more willing to share information because they understand its contribution. Other departments want to participate because they can see tangible results rather than future promises. In complex organisations, ecosystems buy into working solutions, not theoretical slidedecks. Success creates demand, and demand accelerates both intelligence and data maturity.

In this way, intelligence becomes the catalyst for better data. Business value creates clarity. Clarity drives better data. Better data enables more intelligence.



> You don't discover valuable data by organising it. You discover it by solving operational problems.

The engineering principles behind the philosophy.

Making this philosophy work requires an architecture capable of working with fragmented, incomplete and constantly evolving operational environments. Over the last seven years, Entopy has developed four engineering principles that make this approach practical.

Dynamic operational ontology

One of the biggest barriers to AI is data inconsistency. The same operational object may be described differently across multiple systems—a train in one application, rolling stock in another and an asset elsewhere. Traditional approaches attempt to solve this by standardising data across the enterprise, creating a common organisational data model before intelligence can be deployed.

In practice, this is difficult. It requires systems, departments and suppliers to conform to a common organisational data model, often through lengthy transformation programmes that deliver little immediate operational value. Entopy takes a different approach. Rather than standardising the data itself, it standardises the meaning of the data.

A top-level ontology provides a stable, cross-domain operational language. Instead of standardising source systems, it abstracts them into a common set of operational concepts and relationships. Different names, structures and schemas are reduced to the same underlying meaning, allowing intelligence to reason consistently regardless of where the data originated.

Beneath this sits a domain ontology, deployed independently for each customer or operational environment. This captures the terminology, assets, processes and relationships unique to that operation while preserving the common operational language. Combined with a distributed, API-led architecture, each domain evolves independently without affecting any other deployment.

The result is not a single enterprise taxonomy or global data model. It is a shared operational language capable of understanding many different ones.

AI micromodels

Traditional AI often attempts to learn an entire operation as a single problem, requiring vast amounts of data before meaningful intelligence can be created.

Micromodels take the opposite approach. Rather than building one large model, complex operations are decomposed into hundreds of atomic operational questions. Each micromodel answers one question, predicts one behaviour and solves one problem. Smaller questions require less data, are easier to validate and can evolve independently as operational understanding grows.

This changes the deployment model entirely. Each micromodel requires only the data needed to answer a single operational question, allowing organisations to begin creating intelligence long before every dataset has been integrated. As new capabilities are added, new data is introduced where it creates value, operational understanding grows and the intelligence layer evolves incrementally, one capability at a time.

Iterative deployment.

Micromodels and the operational ontology fundamentally change how intelligence is deployed. Rather than waiting for every dataset to be integrated, organisations can begin with a single operational capability, solving one business problem at a time.

Each deployment delivers measurable operational value, builds confidence and reveals where additional data, models and capabilities will have the greatest impact. Future investment is driven by evidence rather than assumption, allowing intelligence to evolve alongside the organisation itself.

This creates a fundamentally different adoption model. Instead of a single, multi-year transformation programme, intelligence grows through a continuous series of operational improvements. Every deployment expands operational understanding, strengthens the data foundation and creates momentum for the next capability.

Synthetic data.

Sometimes the data simply doesn't exist. More often, it exists but is incomplete, sparse or missing the very events organisations want AI to understand. In critical infrastructure, severe weather, major failures and operational disruption occur too infrequently to provide enough training data for traditional machine learning.

Synthetic data addresses this challenge. Rather than replacing operational data, it augments it by generating realistic additional observations that preserve the statistical and operational characteristics of the original dataset. This enables robust models to be developed from sparse and incomplete data, while continuously improving as new operational data becomes available.

Like the broader philosophy described in this paper, synthetic data reduces the need to wait for perfect conditions before creating meaningful intelligence.

Worked example: solving the first problem.

The philosophy of intelligence-led deployment is independent of sector. Whether operating a port, airport, water network or railway, the principle remains the same: begin with a single operational problem rather than attempting to model the entire operation. The goal of a Unified Intelligence layer will come, but the objective is to start somewhere and grow. The below examples are illustrative.

Sector	Started with...	Expanded to...
Port of Dover	Freight vehicle prediction	Tourist traffic → Coaches → Ferry operations → Whole road network intelligence
Harwich Haven Authority / Port of Felixstowe	Tug operations	Vessel ETA → Pilotage → Berth occupancy → Whole-port operational intelligence
Glasgow Airport	Passenger movement prediction	Terminal occupancy → Security demand → Car parks → Whole-airport intelligence
Wastewater (illustrative)	Flood risk prediction	Pump performance → Storage capacity → Network resilience → Catchment-wide intelligence
Electricity Networks (illustrative)	Local demand forecasting	Asset loading → Fault prediction → Maintenance planning → Network-wide intelligence

Conclusion.

The debate should never be whether organisations need better data. They do.

Better governed, better connected and better understood data will always create long-term organisational value. The question is not whether to invest in data. It is when and why.

For many organisations, AI has become trapped behind data transformation. Operational intelligence is delayed until the data estate is considered complete, integrated and AI-ready. In reality, few organisations ever reach that point. Data evolves continuously because operations evolve continuously.

There is another way.

Rather than waiting for perfect data, organisations can begin by solving operational problems. Every successful deployment creates measurable business value. That value identifies which data matters, exposes gaps in quality, encourages wider participation and directs future investment towards the information that demonstrably improves decisions.

In this way, intelligence becomes more than the outcome of better data.

It becomes the mechanism through which better data is created.

This requires a different way of engineering AI: one that embraces fragmented data, learns incrementally and delivers operational capability one problem at a time. Dynamic ontologies, micromodels, iterative deployment and synthetic data are not simply technical innovations. Together, they provide a practical framework for crossing the data chasm without waiting years for perfect foundations.

The organisations that succeed with AI will not necessarily be those with the largest data platforms or the longest transformation programmes. They will be those that create operational value first, allowing intelligence and data maturity to evolve together.

The fastest route to better data may not be waiting for a perfect data foundation.

It may be building better intelligence first.

entopy