

M E R I T

# The Intelligence Equation: Why AI + Human Intelligence is the Only Reliable Formula for Data Harvesting at Scale

How industry-leading platforms combine AI speed and human precision to build data assets their competitors cannot replicate.

Authored by **Daniel Dennis**

Head of Sales - Data Products and Solutions,  
Merit Data and Technology

## CONTENTS

# Table of Contents

### Executive Summary

The Data Quality Crisis: Why Volume is the Wrong Metric

What AI Does Brilliantly: The Case for Automation-First Harvesting

What Humans Do Irreplaceably: The Case for Expert Intelligence

The Intelligence Equation: Designing the Optimal Hybrid Model

The Six Sectors: How the Equation Adapts by Industry

From Data to Competitive Advantage: What the Equation Delivers

### Conclusion

### About Merit Data & Technology

### Sources

# The Intelligence Equation: Key Findings

## SITUATION

Every business that runs on data faces the same challenge. The data that powers competitive advantage cannot be bought off a shelf. It must be harvested - with precision, at scale, and at speed. In 2026, most organisations are using AI to do that harvesting. Most of them are getting it wrong.

## COMPLICATION

The IBM Institute for Business Value found that only 16% of AI initiatives have successfully scaled across the enterprise.<sup>1</sup> S&P Global Market Intelligence reports that 42% of companies abandoned most of their AI initiatives in 2025 - up from just 17% the year before.<sup>2</sup> The data quality crisis is not a technology problem. It is a harvesting problem. And no amount of additional automation solves it.

## RESOLUTION

The organisations whose platforms consistently outperform their markets have found the answer. Not AI alone. Not human teams alone. The two working together - in the right ratio, at the right stages of the data lifecycle. We call this the Intelligence Equation.

## KEY FINDINGS

- 1. Volume is not value.** Over 25% of organisations lose more than \$5 million annually to poor data quality. Organisations harvesting at scale without quality controls are building on sand.
- 2. AI excels at scale, speed, and consistency.** But it fails reliably at contextual interpretation, ambiguity resolution, and domain-specific quality validation - the three things that determine whether data is commercially usable.
- 3. Human expertise is not a cost to minimise.** It is the ingredient that transforms volume into value. The organisations that treat domain expertise as a strategic asset consistently outperform those that treat it as overhead.
- 4. The Intelligence Equation is a five-stage model.** AI-led collection, expert enrichment, AI-powered standardisation, human validation, and continuous intelligent refresh. Each stage has a distinct role. None can be skipped.
- 5. The equation looks different in every sector.** Automotive, Construction, Energy, Healthcare, Maritime, and Retail each have distinct data environments. The principle is universal; the implementation must be bespoke.

## KEY STATISTICS

**16%**

of AI initiatives have successfully scaled across the enterprise  
IBM IBV, 2025

**42%**

of companies abandoned AI initiatives in 2025 - up from 17% in 2024  
S&P Global, 2025

**25%+**

of organisations lose more than \$5M annually due to poor data quality  
Forrester, 2025

# The Data Quality Crisis: Why Volume is the Wrong Metric

## The enterprise data challenge in 2026 is not collection. It is quality.

For a decade, businesses have chased volume - more sources, more records, more signals. The big data era established a working assumption that scale of information would surface insights that smaller datasets could not. That assumption has proven costly.

Forrester research shows that over 25% of organisations lose more than \$5 million annually due to poor data quality - with 7% reporting losses of \$25 million or more.<sup>3</sup> But that figure understates the true commercial impact. Poor data is now the primary reason AI initiatives stall. The IBM Institute for Business Value found that only 29% of technology leaders believe their enterprise data meets the quality standards needed to scale AI.<sup>4</sup> And as organisations push further into generative AI, the problem is intensifying - only 16% of AI initiatives have successfully reached enterprise scale, with data readiness consistently cited as the primary barrier.

**"43% of CDOs cite data quality and readiness as the #1 obstacle to AI success - ahead of cost, skills, or strategy."**

- Informatica CDO Insights Survey, 2025

The root cause is rarely a shortage of data. Organisations are swimming in it. The root cause is how that data is harvested. Automated systems - even sophisticated AI-driven ones - can collect at scale, but they cannot reliably distinguish the accurate from the approximate, the current from the outdated, or the contextually relevant from the superficially matching. Not without human oversight applied at the right points.

## Three Layers of Data Quality Failure

Our data engineering teams have identified three recurring failure points in harvesting operations that rely on automation alone.

**Structural errors.** Unstructured sources - web content, PDFs, regulatory filings, technical specifications - are formatted inconsistently. Automated parsers make systematic errors when data does not conform to expected patterns. Without human review, those errors propagate silently through the dataset.

**Contextual misclassification.** AI systems classify and tag data based on pattern recognition trained on historical examples. When domain-specific terminology is ambiguous - a consistent feature of Energy, Healthcare, and Maritime data environments - misclassification rates climb significantly. Automated systems often proceed with high confidence on a wrong answer.

**Staleness and coverage gaps.** Real-world data changes continuously. Automated systems can be configured to refresh on a schedule, but they cannot assess the significance of a change or prioritise re-harvesting based on commercial importance. Data freshness is not just a frequency problem. It is a judgment problem.

None of these failures are insurmountable. None of them can be resolved by adding more automation. They require human intelligence, applied with domain expertise, at the right point in the harvesting workflow.

## The 1-10-100 Rule of Data Quality

It costs \$1 to verify a record at ingestion. It costs \$10 to correct an error once it has propagated through a pipeline.

It costs \$100 to remediate the impact of an error that reached production. Expert human validation at the harvesting stage is not a luxury - it is the highest-return quality investment in the entire data lifecycle.

# What AI Does Brilliantly: The Case for Automation-First Harvesting

**AI-driven automation has fundamentally changed what is possible in data harvesting. The question is not whether to use it - the answer is unequivocally yes. The question is where.**

Our data collection operations harvest millions of records daily across six industry verticals. That scale is only achievable through AI-driven automation. The competitive question is not whether you are using AI - it is whether your AI is producing data accurate enough, contextually rich enough, and fresh enough to power the decisions your business depends on.

## Where AI Delivers Unambiguous Value

**Scale and speed at low cost.** No human team can match the throughput of a well-configured AI harvesting system. Automated pipelines can ingest, parse, and process millions of records in the time it would take a human analyst to review a few hundred. For industries where competitive intelligence depends on market-wide coverage - Automotive pricing, Energy procurement, Retail catalogue management - this scale is a fundamental requirement, not a nice-to-have.

**Consistency and endurance.** AI systems do not experience fatigue, distraction, or variance in attention. When configured correctly, they apply the same classification logic, the same formatting standards, and the same quality filters to the millionth record as they did to the first. For large-scale, rule-based harvesting tasks, this consistency is a significant advantage over manual processes.

**Real-time monitoring and refresh.** Intelligent data management systems can monitor source changes in near-real-time, triggering re-harvesting workflows when updates are detected. For data products that power live platforms - pricing tools, risk models, market intelligence dashboards - this always-on capability is transformative.

**Pattern recognition at scale.** Modern AI systems can identify relationships, anomalies, and patterns across datasets far too large for human analysis. This capability is increasingly valuable in the enrichment phase of harvesting, where raw records can be enhanced with derived intelligence that adds commercial value.

**76%**

of enterprises now deploy human-in-the-loop processes to validate AI outputs before production  
Enterprise AI Survey, 2025

**47%**

of enterprise AI users made a major business decision based on hallucinated content in 2024  
Enterprise AI Survey, 2025

**43%**

of CDOs cite data quality and readiness as the #1 obstacle to AI success  
Informatica CDO Insights, 2025

## Where Automation Reaches Its Ceiling

The research on human-AI collaboration is instructive. A 2024 meta-analysis published in Nature Human Behaviour, reviewing 108 controlled experiments, found that human-AI combinations often underperformed the stronger of human or AI acting independently - particularly in decision-making tasks where humans anchored too closely to AI outputs.

The lesson is not that human involvement is counterproductive. It is that poorly designed human-AI interaction is counterproductive. The Intelligence Equation is not about adding humans to AI pipelines indiscriminately. It is about inserting human expertise precisely at the points where AI capability reaches its ceiling. In data harvesting, those ceiling points are consistent and identifiable.

# What Humans Do Irreplaceably: The Case for Expert Intelligence

**The most persistent and damaging narrative in enterprise data is that human involvement is a cost to be minimised. It is not. It is the ingredient that determines whether data has commercial value.**

The most sophisticated data operations in the world - those powering platforms that dominate their industries - do not treat human expertise as a residual cost. They treat it as a strategic asset: the ingredient that transforms volume into value, and automation into competitive advantage.

**"This foundation of structured, accessible, high-quality data represents the essential precondition for sustained AI success."**

- IBM Institute for Business Value, 2025

## Four Domains Where Human Intelligence is Irreplaceable

**Domain expertise and contextual interpretation.** A Construction sector analyst understands what a change in planning application terminology signals about project pipeline. An Energy specialist can interpret the commercial significance of a regulatory amendment that an AI system would process as routine text. This contextual layer - the ability to understand what data means, not just what it says - is the difference between a data product and a data asset.

**Ambiguity resolution.** Real-world data is inherently ambiguous. The same term means different things in different contexts. A product category clearly defined in one market is fluid in another. Regulatory language precise in its original form becomes ambiguous across jurisdictions. AI systems handle ambiguity through probabilistic inference - they make their best guess. Human experts resolve ambiguity through knowledge - they know the right answer.

**Quality validation and exception handling.** Applying the 1-10-100 rule: it costs \$1 to verify a record at ingestion, \$10 to correct it downstream, and \$100 to remediate the impact of an error in production.<sup>7</sup> Skilled data analysts applying validation logic at the point of harvesting are not a cost - they are the most efficient quality investment in the pipeline.

**Strategic data architecture.** The structure of a dataset - how records are categorised, what attributes are captured, how relationships are modelled - determines its downstream usefulness for years. These decisions require human expertise in both the domain and the intended use case. A dataset architected well at the start compounds in value. One architected poorly is expensive to fix and commercially expensive to use.

## Why Sector Expertise Cannot Be Automated

One of the most consistent findings from our work across Automotive, Construction, Energy, Healthcare, Maritime, and Retail is that no two industry data environments are alike. Each sector has its own terminology, its own regulatory framework, its own data conventions, and its own definition of what constitutes a high-quality record.

Cross-sector AI models that promise a universal solution to data harvesting consistently underperform bespoke approaches because they cannot be simultaneously expert in all these environments. The IBM IBV found that 49% of executives cite data inaccuracies and bias as a primary barrier to adopting agentic AI.<sup>8</sup> In almost every case, that poor quality originates in the harvesting stage - and it originates because the systems doing the harvesting lack the domain knowledge to know when they are wrong.

### Merit's Domain Expertise Model

Our sector specialists are not editors reviewing AI output for errors. They are the interpretive layer that transforms raw harvested data into sector-specific intelligence. This is not a supplementary feature of how we work. It is the core of it.

# The Intelligence Equation: Designing the Optimal Hybrid Model

**The Intelligence Equation is not a fixed ratio of AI to human involvement. It is a design principle: match AI capability and human expertise to the specific requirements of each stage of the data lifecycle.**

What this looks like in practice depends on data type, required accuracy threshold, volume, and the downstream use case. But there are consistent patterns that characterise best-practice implementations - patterns we have developed and refined across six sectors and hundreds of bespoke dataset builds.

## Stage 1 - AI-Led Collection at Scale

The collection phase is where AI-driven automation delivers its clearest value. The priorities are coverage, speed, and consistency. AI systems monitor thousands of sources simultaneously, identify new records, and ingest them into structured pipelines at a scale no human team could match. Our AI-driven infrastructure processes millions of data points daily across global sources in six industry verticals. Human involvement at this stage is limited to system configuration, source selection strategy, and monitoring for collection anomalies.

## Stage 2 - Expert Research and Enrichment

Raw collected data is rarely usable in its harvested form. It requires enrichment - the addition of context, relationships, and derived intelligence that transforms a record from data point to data asset. This is where human expertise delivers its highest return. Our sector specialists conduct primary research to validate and enrich data that automated systems cannot access or interpret reliably: verification of technical specifications against authoritative sources, interpretation of regulatory changes and their commercial implications, mapping of entity relationships, and assessment of data significance in the context of current market conditions.

## Stage 3 - AI-Powered Standardisation and Processing

Once enriched, data must be standardised and processed into the formats downstream systems require. This is a high-volume, rule-based task well suited to AI automation. Natural language processing models transform unstructured text into structured attributes. Classification algorithms apply consistent taxonomy at scale. The critical point is that the rules being applied have been defined and validated by human experts. The AI is executing with precision; the intelligence governing what precision looks like came from domain expertise invested in stages one and two.

## Stage 4 - Human Validation and Quality Assurance

Before any dataset enters production, human validation is applied to ensure the full pipeline has performed to the required accuracy threshold. This is not a wholesale manual review of every record. It is statistically designed sampling, exception-based review, and domain-expert sign-off on the quality metrics that matter for each use case. Errors caught here cost a fraction of errors caught downstream. And errors that reach a production pricing algorithm, risk model, or AI training set can have consequences that far exceed the cost of the dataset itself.

## Stage 5 - Continuous Intelligence and Refresh

Data harvesting is not a point-in-time activity. The Intelligence Equation is an ongoing operational model - continuously monitoring for source changes, prioritising re-harvesting based on commercial significance, and applying the same AI-and-human quality standards to updates as to the original build. This always-on, always-fresh model is what distinguishes a data partner from a data vendor. A vendor delivers a dataset. A partner maintains the intelligence that powers your decisions.

**Millions**

of data points harvested daily across six industry verticals by Merit's AI-driven pipelines

**6**

sector domains - Automotive, Construction, Energy, Healthcare, Maritime, Retail - each requiring bespoke domain expertise

# The Six Sectors: How the Equation Adapts by Industry

---

**The Intelligence Equation looks different in every sector. The underlying principle is consistent - AI-driven scale combined with human expertise. But the balance, the critical intervention points, and the quality requirements are determined by each sector's distinct data environment.**

## Automotive - Speed and Precision in Pricing Intelligence

Automotive data moves fast. Vehicle specifications, pricing, inventory, and market positioning can shift materially within hours in active market conditions. AI-driven collection and near-real-time refresh are non-negotiable. But the interpretation of pricing signals - understanding why a price has moved, whether a specification change is commercially significant, or how a new model launch reshapes competitive positioning - requires domain expertise that automated systems cannot replicate.

## Construction - Complexity and Regulatory Navigation

Construction sector data - planning applications, project specifications, regulatory approvals, contractor intelligence - is notoriously unstructured and jurisdiction-specific. AI systems can harvest at volume, but the interpretation of planning documents, the identification of commercially significant project milestones, and the navigation of regulatory variance across jurisdictions require expert human analysis. The Intelligence Equation in Construction is weighted more heavily toward human expertise than almost any other sector.

## Energy - Market Intelligence in a Volatile Environment

Energy market data operates at the intersection of commodity pricing, regulatory compliance, and geopolitical intelligence. The speed requirements favour AI-driven collection; the complexity requirements demand expert interpretation. Pricing data alone is insufficient - the intelligence that powers Energy sector platforms combines price signals with regulatory analysis, supply chain intelligence, and forward-looking market context. This multi-dimensional dataset can only be built through AI scale paired with human expertise.

## Healthcare - Accuracy as a Non-Negotiable

In Healthcare data - formularies, clinical intelligence, regulatory submissions, product specifications - accuracy is not a quality metric. It is a regulatory and ethical requirement. The error tolerance that might be acceptable in a Retail catalogue dataset is simply not available in Healthcare. The Intelligence Equation in Healthcare is structured accordingly: AI-driven collection provides coverage at scale, but human expert validation at multiple pipeline stages ensures the accuracy standards required for clinical and commercial use are consistently met.

## Maritime - Global Coverage with Local Precision

Maritime data - vessel specifications, fleet intelligence, port activity, regulatory compliance - requires global coverage at a scale that only AI can deliver. But interpretation of vessel data requires specialist knowledge of classification standards, flag state regulations, and operational characteristics that vary significantly across regions and vessel types. Human experts with Maritime domain knowledge are essential to the enrichment and validation stages.

## Retail - Volume, Freshness, and Attribute Completeness

Retail data combines three simultaneous challenges: the scale challenge (millions of SKUs across thousands of sources), the freshness challenge (prices and availability change continuously), and the completeness challenge (the specific attributes that make a record useful vary by category and use case). AI-driven automation handles collection and standardisation; human expertise defines the attribute taxonomy, validates completeness against category requirements, and ensures the data structure serves the platforms it is designed to power.

# From Data to Competitive Advantage: What the Equation Delivers

**The purpose of the Intelligence Equation is not data for its own sake. It is the competitive advantage that high-quality, sector-specific, AI-harvested and human-validated data enables.**

## AI-Readiness - The Data Foundation That Determines AI Success

The most consequential application of high-quality harvested data in 2026 is as training and grounding data for AI systems. Enterprise AI adoption has created enormous demand for data that meets AI-ready standards - accurate, structured, comprehensive, and domain-specific. The IBM IBV is clear: only 16% of AI initiatives have successfully scaled across the enterprise, and data readiness is consistently the differentiating factor.<sup>9</sup> The organisations that win the AI race are not those with the most advanced models.

They are those with the most accurate, relevant, and comprehensive data to power those models. That data cannot be harvested by automation alone.

**"68% of AI-first organisations report mature, well-established data and governance frameworks - compared with just 32% of other organisations. The data foundation is the differentiator."**

- IBM Institute for Business Value, 2025

## Competitive Moat - Data Your Competitors Cannot Replicate

Generic, off-the-shelf data is available to every organisation in a sector. It provides no competitive advantage because it provides no differentiation. Deloitte's State of AI in the Enterprise 2026 found that 42% of companies abandoned AI initiatives where data quality issues proved insurmountable.<sup>10</sup> The organisations whose platforms consistently outperform are those whose data assets cannot be purchased by a competitor - because they were built, not bought. Bespoke datasets built through the Intelligence Equation encode sector expertise, source relationships, enrichment logic, and quality standards that compound in value over time. This is the data moat.

## Decision Velocity - From Data to Action, Faster

High-quality, real-time harvested data shortens the distance between information and action. It reduces due diligence burden because the data can be trusted. It accelerates modelling cycles because the inputs are clean. And it improves the quality of decisions because the intelligence is more complete, more current, and more contextually rich. In sectors where competitive position is determined by the speed and accuracy of commercial decisions - pricing, procurement, market entry, risk assessment - data quality is not an operational issue. It is a strategic one.

## Scalability - Data Operations That Grow with the Business

One of the most commercially significant advantages of AI-driven harvesting infrastructure is its scalability. Deloitte's State of AI in the Enterprise 2026 reports that worker access to AI rose 50% in 2025, with the number of companies scaling AI in production set to double within six months.<sup>11</sup> Once collection pipelines, enrichment logic, and quality standards are established, expanding coverage - new sources, new geographies, new data types - is a function of configuration, not headcount. AI provides the throughput. Human expertise provides the quality governance. Together, they create a data operation whose capacity grows without a linear increase in cost.

## Conclusion

**The data quality crisis is real, measurable, and costing organisations billions. The solution is not more automation. It is not more human analysts. It is the correct combination of both.**

The Intelligence Equation - AI-driven collection at scale, human expertise for enrichment and contextual interpretation, AI-powered standardisation, human validation for quality assurance, and continuous intelligent refresh - is the only harvesting model that consistently produces data at the accuracy, freshness, and relevance that enterprise AI systems and business-critical platforms demand.

For organisations operating in Automotive, Construction, Energy, Healthcare, Maritime, Retail, or any sector where data quality determines competitive position, the question is not whether to invest in the Intelligence Equation. It is how quickly you can implement it before your competitors do.

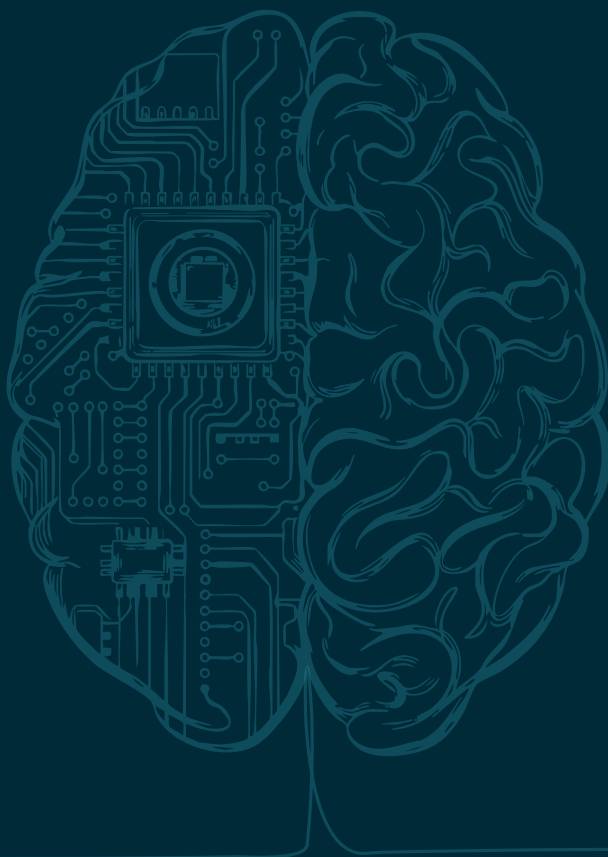
**Ready to build a data asset your competitors cannot replicate?**

Tell us about your data challenge - whatever the industry, whatever the scale.

[www.meritdata-tech.com/industry-data](http://www.meritdata-tech.com/industry-data)

## About Merit Data & Technology

Merit Data & Technology, part of Merit Group PLC, is a trusted partner in AI-driven data and digital transformation. With over two decades of experience, we deliver scalable, secure and AI-ready automation and data solutions tailored to business needs.



## Sources

1. IBM Institute for Business Value (2025). 2025 CEO Study: 5 mindshifts to supercharge business growth. [ibm.com/thought-leadership/institute-business-value](https://ibm.com/thought-leadership/institute-business-value)
2. S&P Global Market Intelligence (2025). Enterprise AI Adoption Survey. Survey of 1,000+ enterprises across North America and Europe. [spglobal.com](https://spglobal.com)
3. Forrester Research (2025). The State of Data Quality. Over 25% of organisations report losses exceeding \$5M annually due to poor data quality; 7% report losses of \$25M+. [forrester.com](https://forrester.com)
4. IBM Institute for Business Value (2024). AI-Ready Data Survey. Only 29% of technology leaders strongly agree their enterprise data meets quality, accessibility, and security standards needed to scale generative AI. [ibm.com/thought-leadership/institute-business-value](https://ibm.com/thought-leadership/institute-business-value)
5. IBM Institute for Business Value (2025). 2025 CEO Study. Only 16% of AI initiatives have successfully scaled across the enterprise. [ibm.com/thought-leadership/institute-business-value](https://ibm.com/thought-leadership/institute-business-value)
6. Vaccaro, M., Almaatouq, A. & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, October 2024.
7. The 1-10-100 rule is a widely cited data quality principle. See also: Acceldata (2026). Turn Data Quality Risks Into Revenue with ADM.
8. IBM Institute for Business Value (2025). AI at the Core: From AI Projects to Profits. 49% of executives cite data inaccuracies and bias as a barrier to adopting agentic AI. [ibm.com/thought-leadership/institute-business-value](https://ibm.com/thought-leadership/institute-business-value)
9. IBM Institute for Business Value (2025). 2025 CEO Study. AI scaling analysis across global enterprise leaders. [ibm.com/thought-leadership/institute-business-value](https://ibm.com/thought-leadership/institute-business-value)
10. Deloitte (2026). State of AI in the Enterprise 2026. Survey of 3,235 senior leaders across 24 countries, August-September 2025. [deloitte.com/global/en/issues/generative-ai/state-of-ai-in-enterprise.html](https://deloitte.com/global/en/issues/generative-ai/state-of-ai-in-enterprise.html)
11. Deloitte (2025). State of AI in the Enterprise 2026. Based on a global survey of 3,235 senior leaders across 24 countries, August-September 2025.