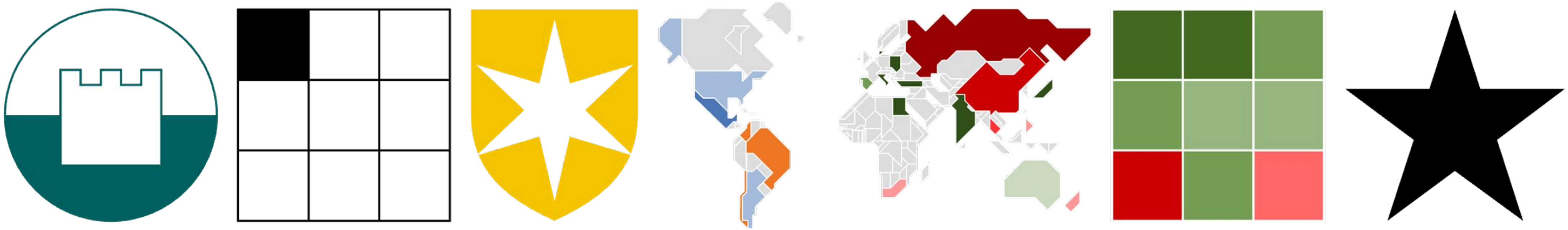
Building Trust in the AI Era of Enterprise Analytics: From Data Integrity to Insight Acceleration

Anusha Dwivedula

Aug 2025



We are the global leader for investor-first insights, tools, and investment strategies, trusted by market participants worldwide. Our system of capabilities, connected through data and insights, is driven by our mission to empower investor success.

Shared language put to work across our broad reach

Individual Investors

4.2M individual investors

3.7M retirement plan participants

Financial Advisors

300,000 financial advisors

Debt Issuers

4,100 debt issuers

Retirement Plan Providers

319,000 retirement plan sponsors

Asset Managers

3,500 asset management firms

Private Market Participants

112,000 PitchBook users

INFLUENCERS

Fintechs & Redistributors

925 global alliances

Regulators

70 regulators globally

Media Companies

130 media companies

Data as of Q3 2024

Recent AI Trust Failures

Why Delta Air Lines Is Facing Backlash for AI Pricing

Aug 06, 2025

Airlines Delta Air Lines Opinion Technology Travel News

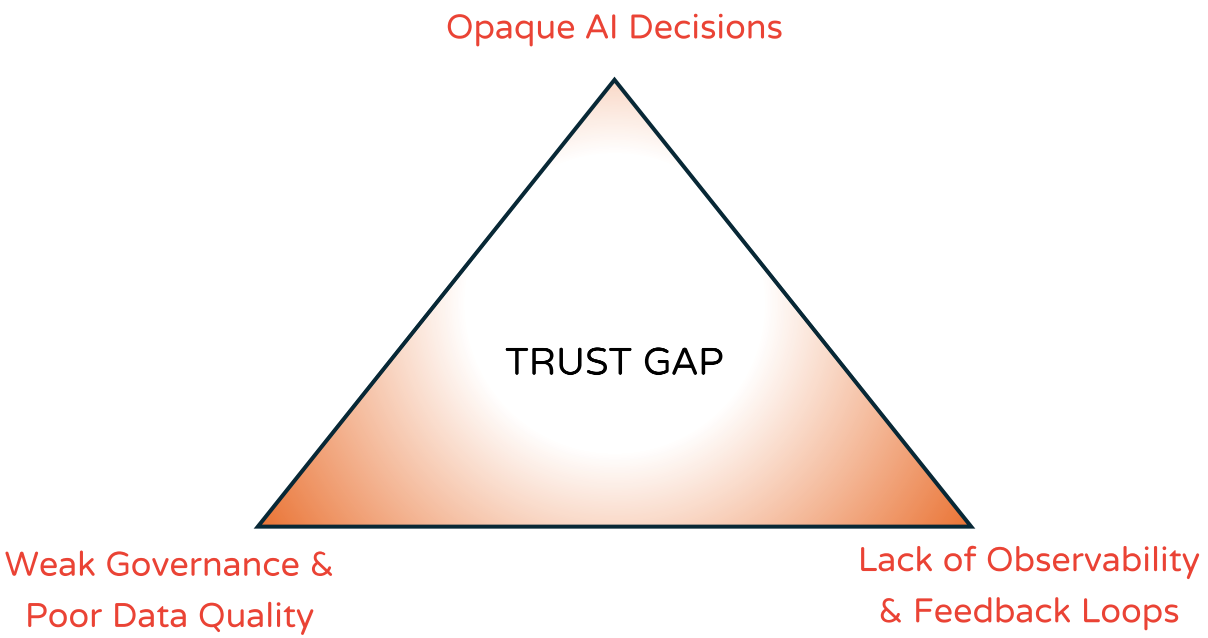
INDEPENDENT

New AI tool picks up every minor scratch on your car rental – and its freaking people out

James Liddell
Thu 10 July 2025 at 11:05 am GMT-5
3 min read

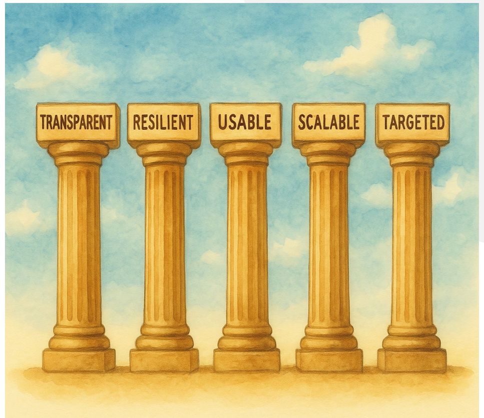
MORNINGSTAR

The Common Thread - The Thing Nobody Talks About Until It’s Gone



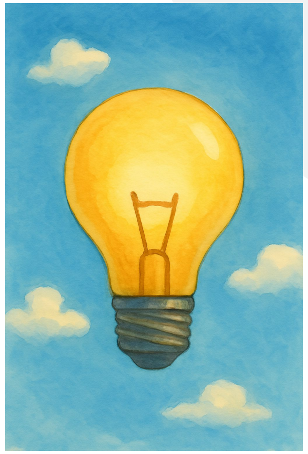
How do you Build Trust in the AI era of Enterprise Analytics?

AI Driven Quality System



The T.R.U.S.T Pillars

Trust Sandwich Framework



How to Enforce?

Morningstar Total Return Index

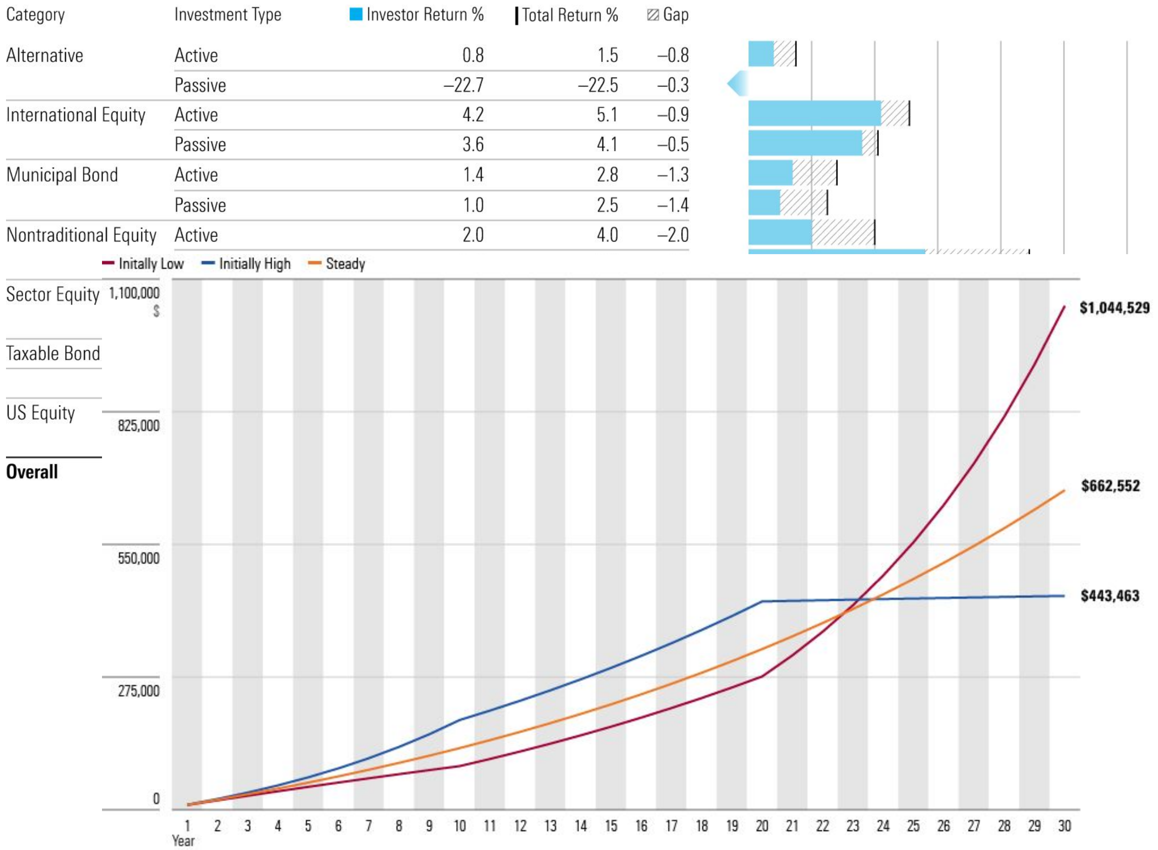
- Total return of a fund by tracking the change in net asset value (NAV), reinvesting all income and capital gains distributions, and dividing by the starting NAV.

- Measure the performance of a fund over time, including both price appreciation and the reinvestment of dividends and distributions

- Various applications – Fund and Category Performance, Benchmarking and Investable Products.

- Errors can break client confidence or create regulatory risk.

- Consumers expect to get access to the data within 15 minutes after market

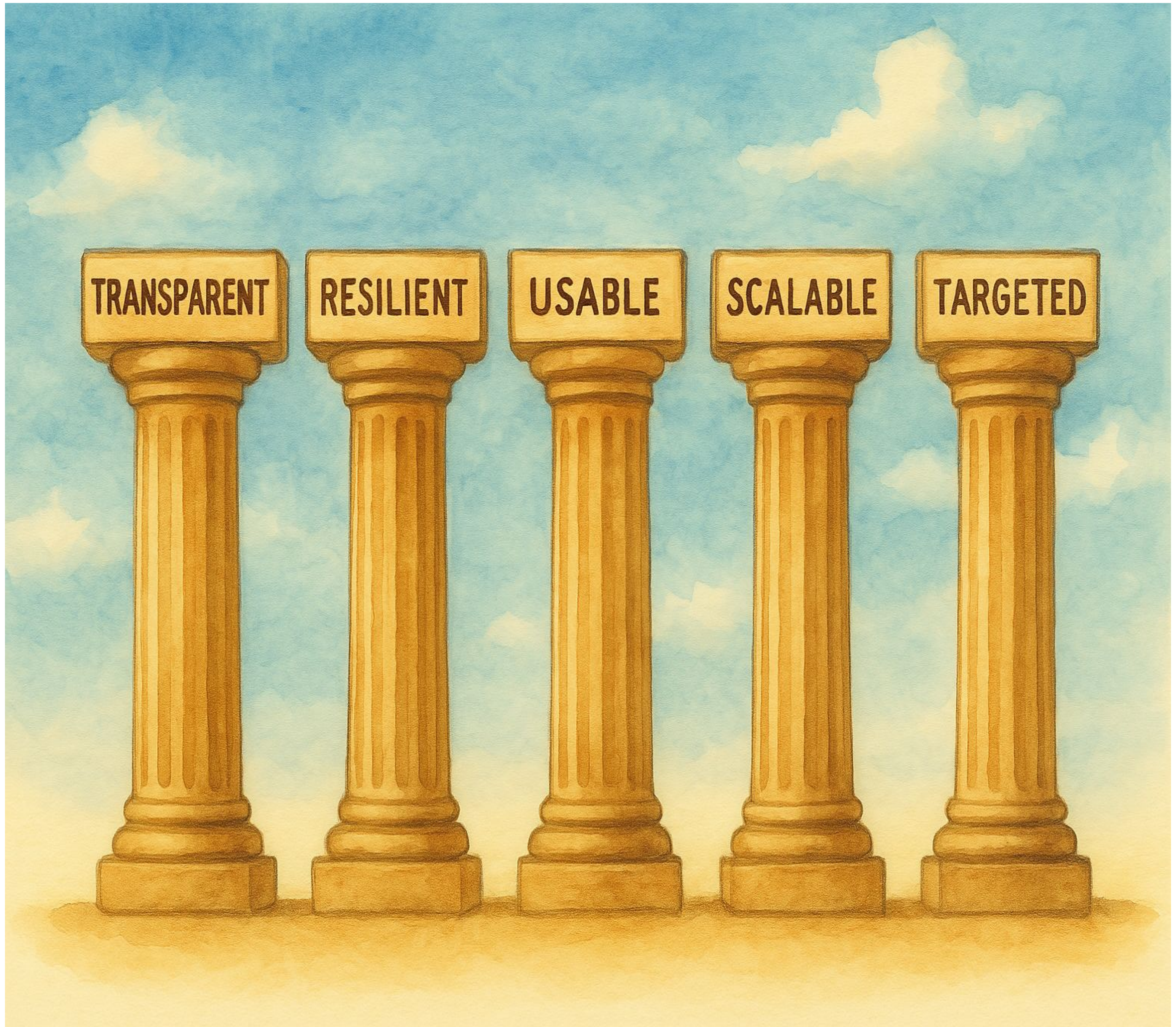


Category	Investment Type	Investor Return %	Total Return %	Gap
Alternative	Active	0.8	1.5	–0.8
	Passive	–22.7	–22.5	–0.3
International Equity	Active	4.2	5.1	–0.9
	Passive	3.6	4.1	–0.5
Municipal Bond	Active	1.4	2.8	–1.3
	Passive	1.0	2.5	–1.4
Nontraditional Equity	Active	2.0	4.0	–2.0
Sector Equity				
Taxable Bond				
US Equity				
Overall				

AI-driven quality system

"We needed a system that could be accurate, fast, and trusted—at scale."

The T.R.U.S.T five pillars of confidence in production systems



The T.R.U.S.T five pillars of confidence in production systems

Transparent — You can trace every decision and every datapoint.

Resilient — It adapts to change and handles uncertainty without breaking.

Usable — Alerts and insights are clear, actionable, and aligned with how teams work.

Scalable — It can handle large, complex, fast-changing systems without slowing down.

Targeted — The system focuses on what matters—no noise, just meaningful signals.

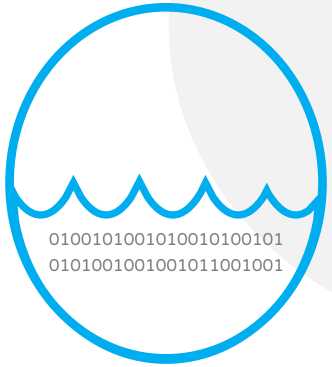
MORNINGSTAR

Introducing the TRUST sandwich framework

- Trust isn't built in one place.
- It's a layered system—like a sandwich—where infrastructure, logic, and human validation work together.

Layer 1: Data Infrastructure and Governance (Transparent and Scalable)

Enterprise Data Platform Journey

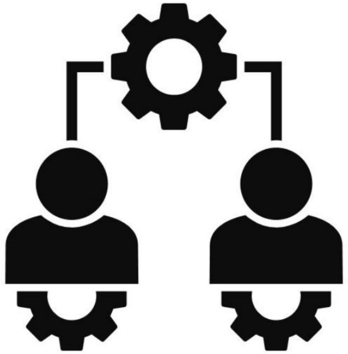


Centralized Data Lake and Lakehouse

2018

Self-Service Platform

2021

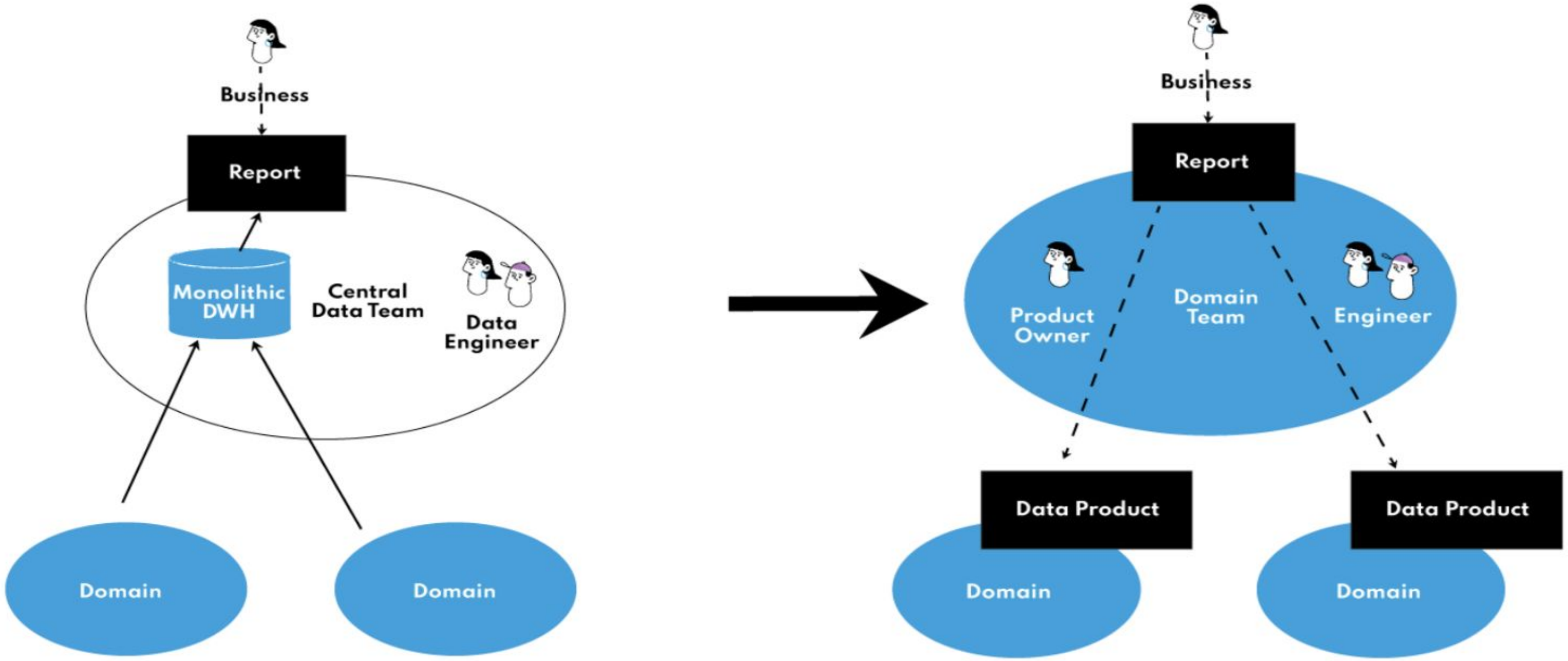


Data Mesh Operating Model

2024 – present

Layer 1: Data Infrastructure and Governance (Transparent and Scalable)

De-centralized Sociotechnical Approach

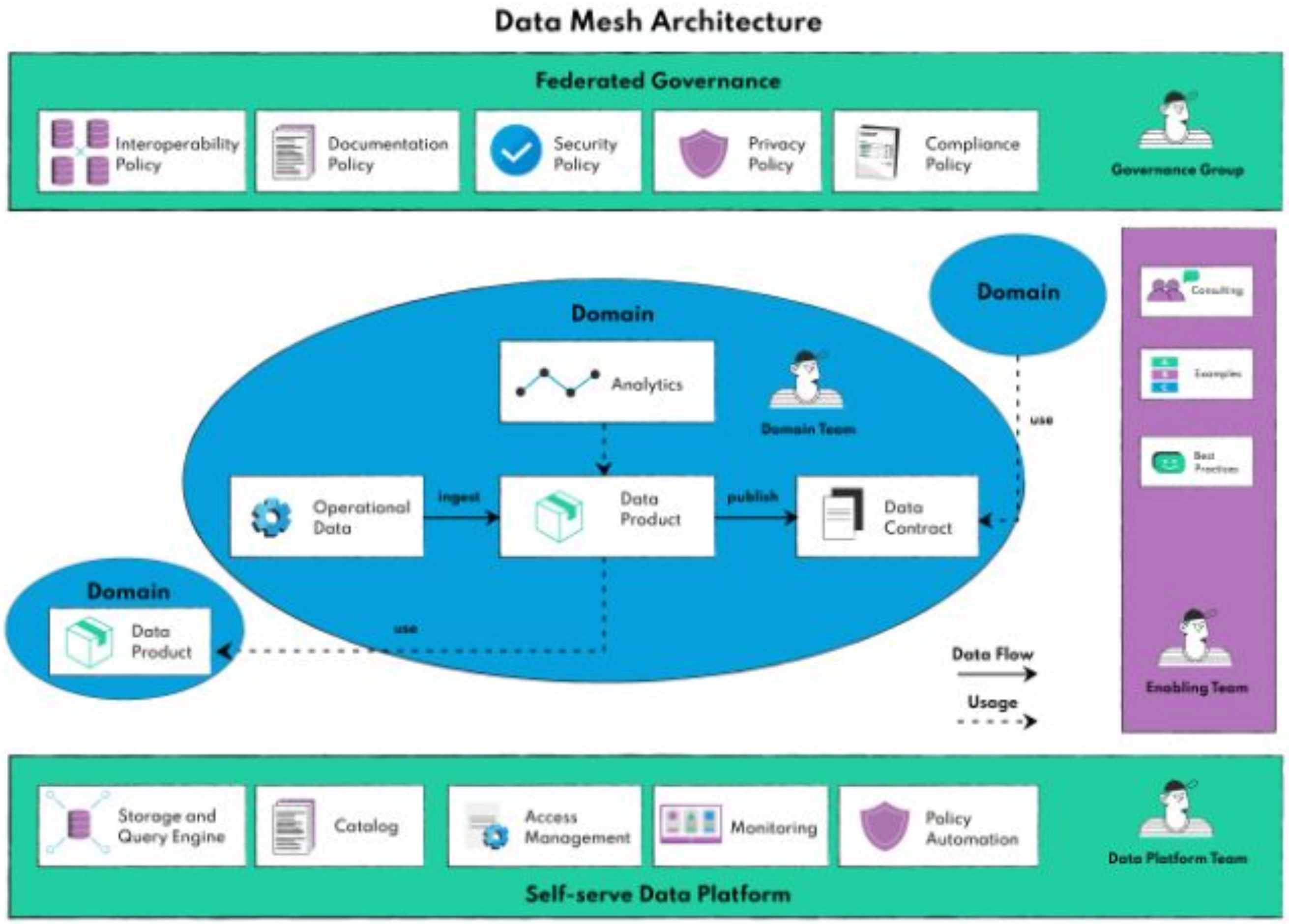


"Data mesh is a decentralized sociotechnical approach to share, access, and manage analytical data in complex and large-scale environments—within or across organizations"

\- Dehghani, Zhamak

Layer 1: Data Infrastructure and Governance (Transparent and Scalable)

Data as a Product



Layer 2: AI-driven detection logic and monitoring (Usable and Targeted)

Monitoring Volume of Daily Ingested Data

Algorithm 3 LSTM-Based Data Volume Monitoring

Input: Historical data volumes $V \in \mathbb{R}^n$, external features $X \in \mathbb{R}^{n \times m}$, lag length l
Normalize V and X using Min-Max scaling
Prepare lagged sequences of length l for V and X
Define and compile an LSTM model using TensorFlow/Keras
Train the model on the lagged sequences
Predict expected volume $\hat{v}_t$ for today
Compute residual: $r_t = v_t - \hat{v}_t$
Compute the 95th percentile threshold: threshold $=$ $\text{Percentile}_{95}(r_t)$
if $|r_t| >$ threshold **then**
    Flag t as an anomaly
end if
Output: Prediction $\hat{v}_t$, anomaly status

LSTM is the preferred approach to monitor volume given the need for scalability across different tables*

Algorithm 4 SARIMAX-Based Data Volume Monitoring

Input: Historical data volumes $V \in \mathbb{R}^n$, external features for calendar effect $X \in \mathbb{R}^{n \times m}$, SARIMAX parameters (p, d, q, P, D, Q, s)
Fit a SARIMAX model: $V_t = \phi(V_{t-1}, ..., V_{t-p}, X_t, ..., X_{t-m}) + \epsilon_t$
Forecast expected volume $\hat{v}_t$
Compute residual: $r_t = v_t - \hat{v}_t$
Compute the 95th percentile threshold: threshold $=$ $\text{Percentile}_{95}(r_t)$
if $|r_t| >$ threshold **then**
    Flag t as an anomaly
end if
Output: Prediction $\hat{v}_t$, anomaly status

SARIMAX is suitable for cases where interpretability and lower computational cost*

*Algorithms cited from IEEE paper we published

Layer 3: Human Context and Expert Validation (Resilient)

Dependency Aware Detection and Customizable Alerts

Algorithm 5 LSTM-Based Anomaly Detection with Expert Opinion

```
Input: TRI series Ψ ∈ ℝ^n, lag length l, threshold θ, expert validation set E
Normalize Ψ using Min-Max scaling
Prepare lagged sequences of length l
Define and train an LSTM model on historical TRI data
for each time step t in 1, ..., n do
    Predict expected TRI: Ψ̂_t = f(Ψ_{t−1}, ..., Ψ_{t−l})
    Compute residual: r_t = Ψ_t − Ψ̂_t
    if |r_t| > θ then
        if E_t = 1 (expert confirms valid data) then
            Suppress false alarm
        else
            Flag t as an anomaly
        end if
    end if
end for
Output: List of detected anomalies with expert corrections
```

Expert-guided AI quality control, unnecessary alerts are reduced, increasing the system's efficiency*

Algorithm 6 Dependency-Aware Anomaly Detection in TRI

```
Input: TRI series Ψ ∈ ℝ^{n×m}, sector mapping 𝓘, threshold θ, sector-wide threshold τ
Normalize Ψ using Min-Max scaling
Group investments into sectors using predefined mappings
for each time step t in 1, ..., n do
    for each sector s ∈ 𝓢 do
        Compute sector-wide return S^s_t
        for each investment i ∈ s do
            Predict expected TRI: Ψ̂t^i = f(Ψt − 1^i, ..., Ψ^i_{t−l})
            Compute residual: r^i_t = Ψ^i_t − Ψ̂t^i
            if |r^i_t| > θ and |S^s_t − S^s_{t−1}| ≤ τ then
                Flag i at time t as an anomaly
            end if
        end for
    end for
end for
Output: List of detected anomalies
```

Reducing False Alarms by Analyzing Data Dependencies*

Algorithm 7 Customizable Alerting System for TRI Anomalies

```
Input: TRI series Ψ, predicted values Ψ̂, user-defined severity θ_s, persistence level p, frequency threshold τ
Initialize breach counter B_w = 0
for each time step t in 1, ..., n do
    Compute residual: r_t = Ψ_t − Ψ̂_t
    if |r_t| > θ_s then
        Increment counter B_w = B_w + 1
        if breach persists for p consecutive steps then
            Trigger persistence-based alert
        end if
    end if
    if B_w > μ_B + τ then
        Trigger breach frequency alert
    end if
end for
Output: List of triggered alerts based on user preferences
```

Avoiding Alert Fatigue Through Customizable Alerting Mechanisms*

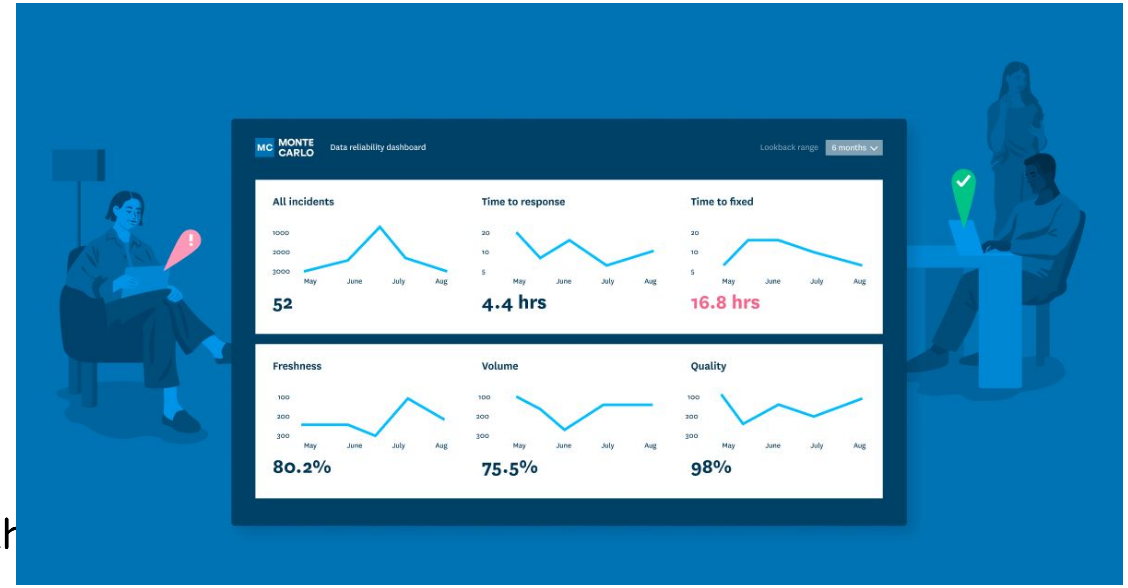
Goal: Alert when it matters!

*Algorithms cited from IEEE paper we published

Layer 3: Human Context and Expert Validation (Resilient)

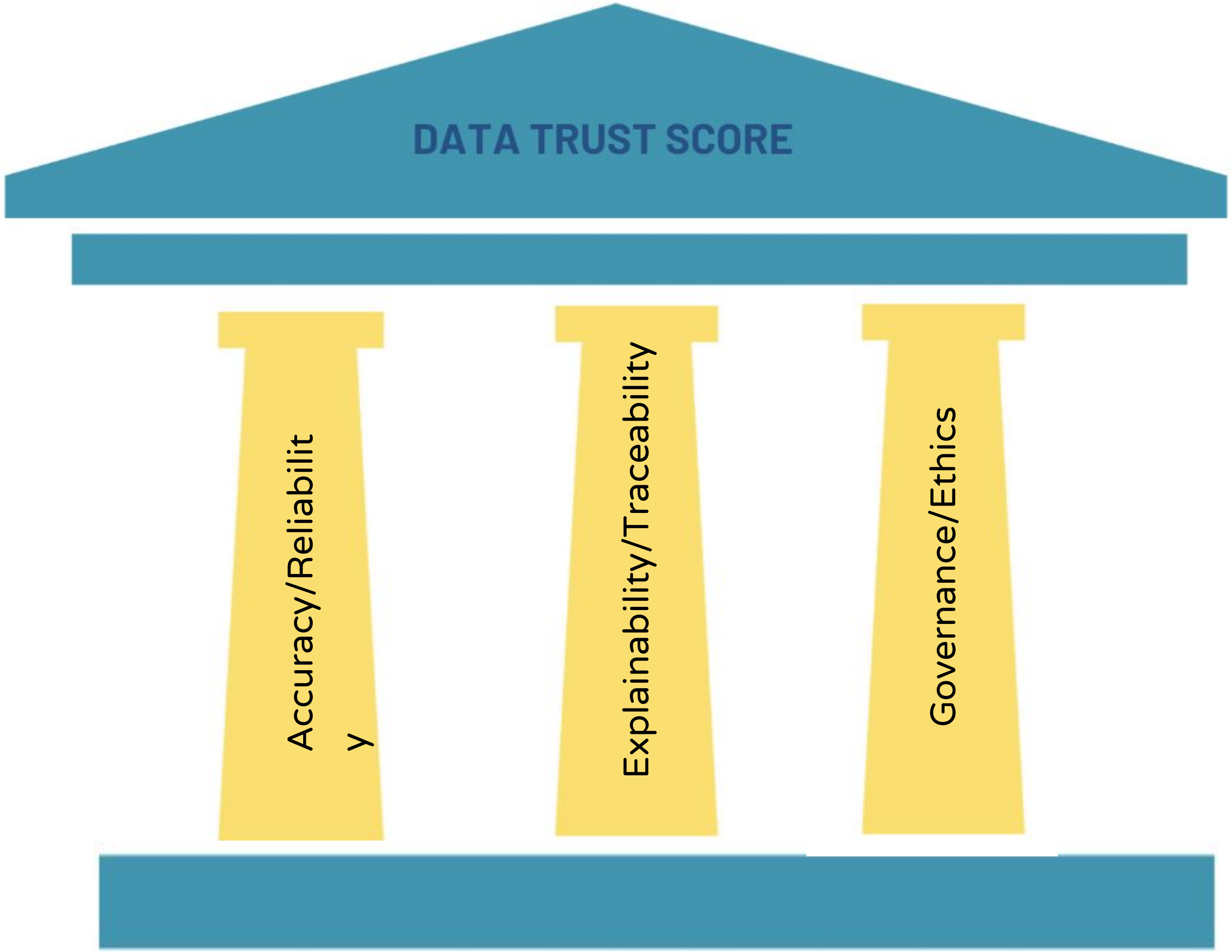
Data Observability

- Experts validate anomalies flagged by the system.
- Confirmed expected fluctuations (e.g., known market patterns)
- Built a feedback loop to refine detection logic over time.
- This gave end-users confidence that the system aligned with reality.



Monte Carlo Tool

Data Trust Score



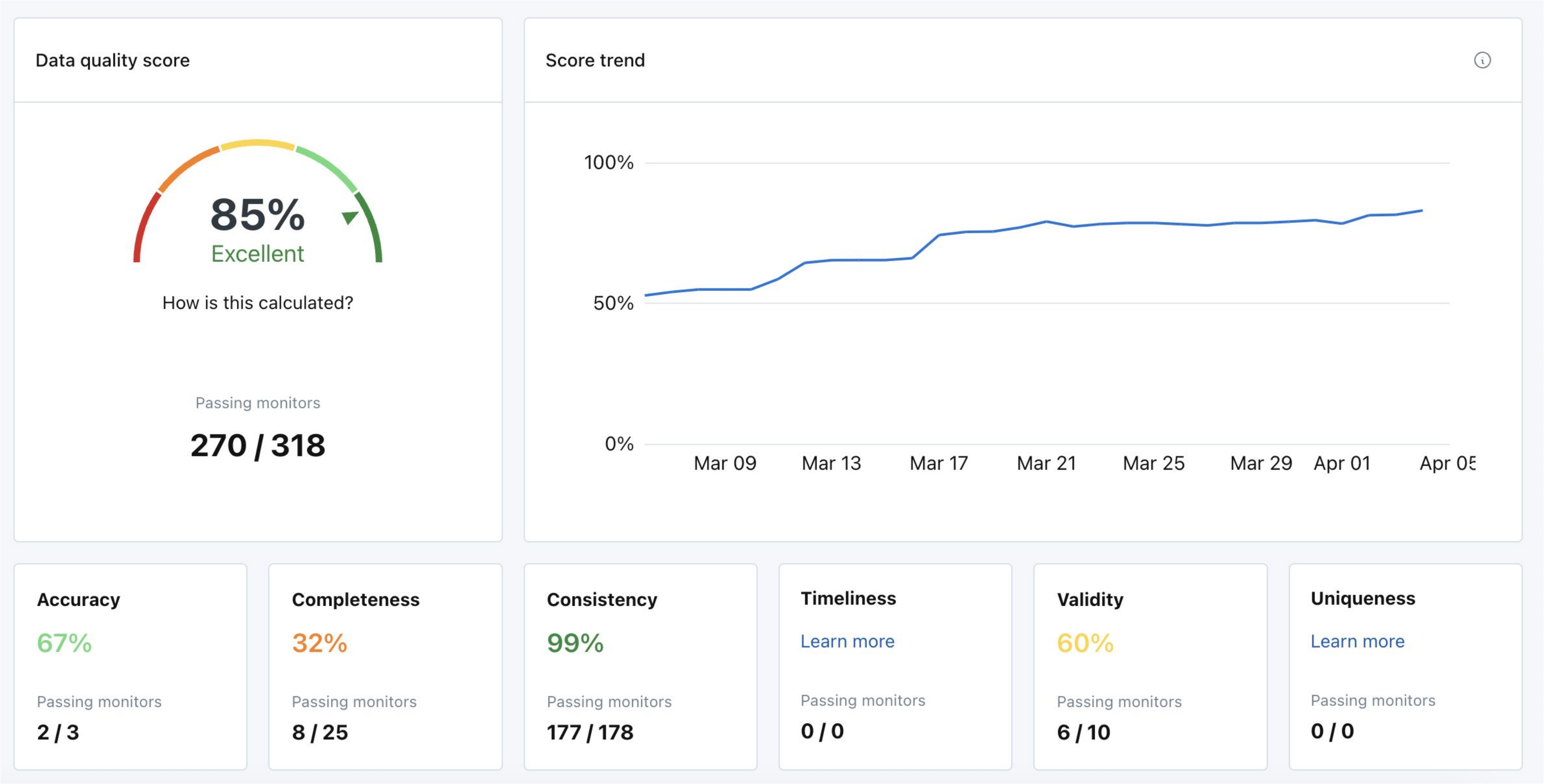
Algorithm 4 Final Data Trust Score Calculation

1: Calculate ARS, ETS, and ECGS as described.
2: Aggregate the pillar scores:
3: $DTS = w_1 \times \text{ARS} + w_2 \times \text{ETS} + w_3 \times \text{ECGS}$

where w_1, w_2, and w_3 are adjustable weights that can be customized based on industry requirements.

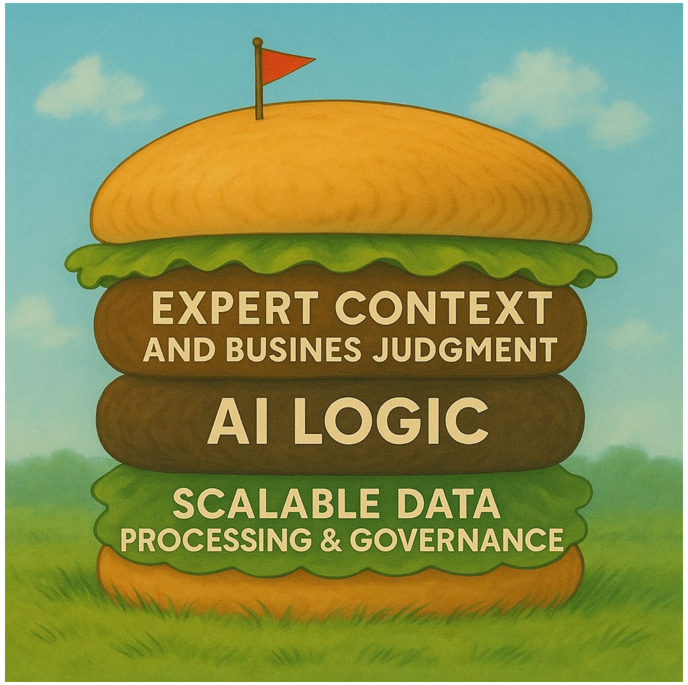
Results

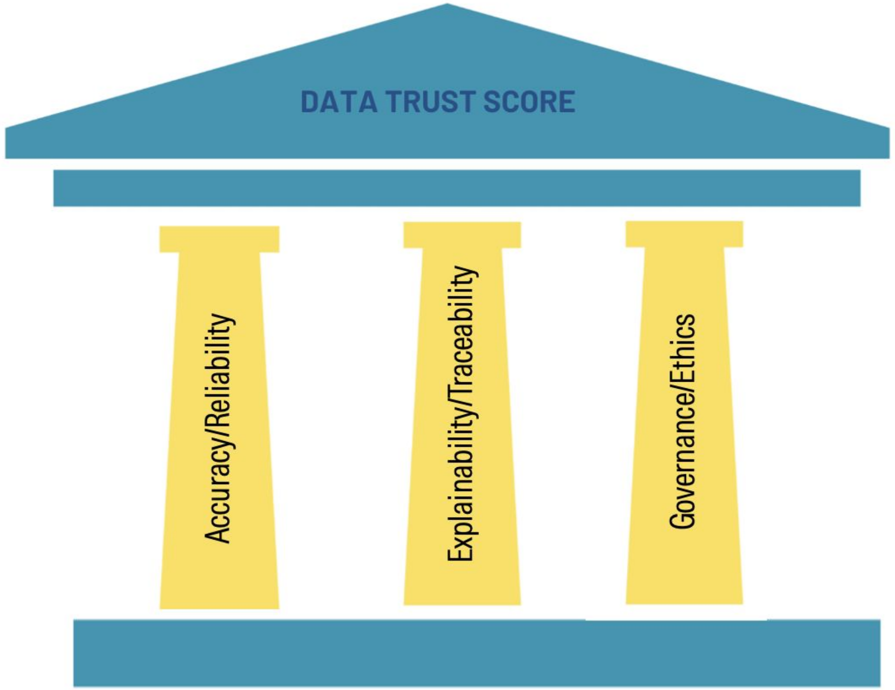
Increased Confidence In Data



“Where is your system leaking trust today?”

Trust is not a feature—it’s a design decision. And it must be built in layers.





Thank You!

- Establishing Trust in AI-Driven Data Oberservability and Quality Control: A Framework for Reliable and Scalable Standards
- Trust Erosion: 4 Signals Your Data Strategy Is Breaking Down
(Before AI Fails)
- Personal Website