WHY DATA ENGINEERING IS THE BACKBONE OF AN AI-FIRST ENTERPRISE

Table of Contents

01. AI Is Only as Good as the Data Behind It 4
02. The Relationship Between Data Quality and AI Outcomes 5
03. Data Engineering Enables Scalable AI 6
04. Data Engineering at Enterprise Scale 7
05. Why Data Engineering Is Non-Negotiable for AI-First Enterprises 7
06. Conclusion 8
07. About the Author 8

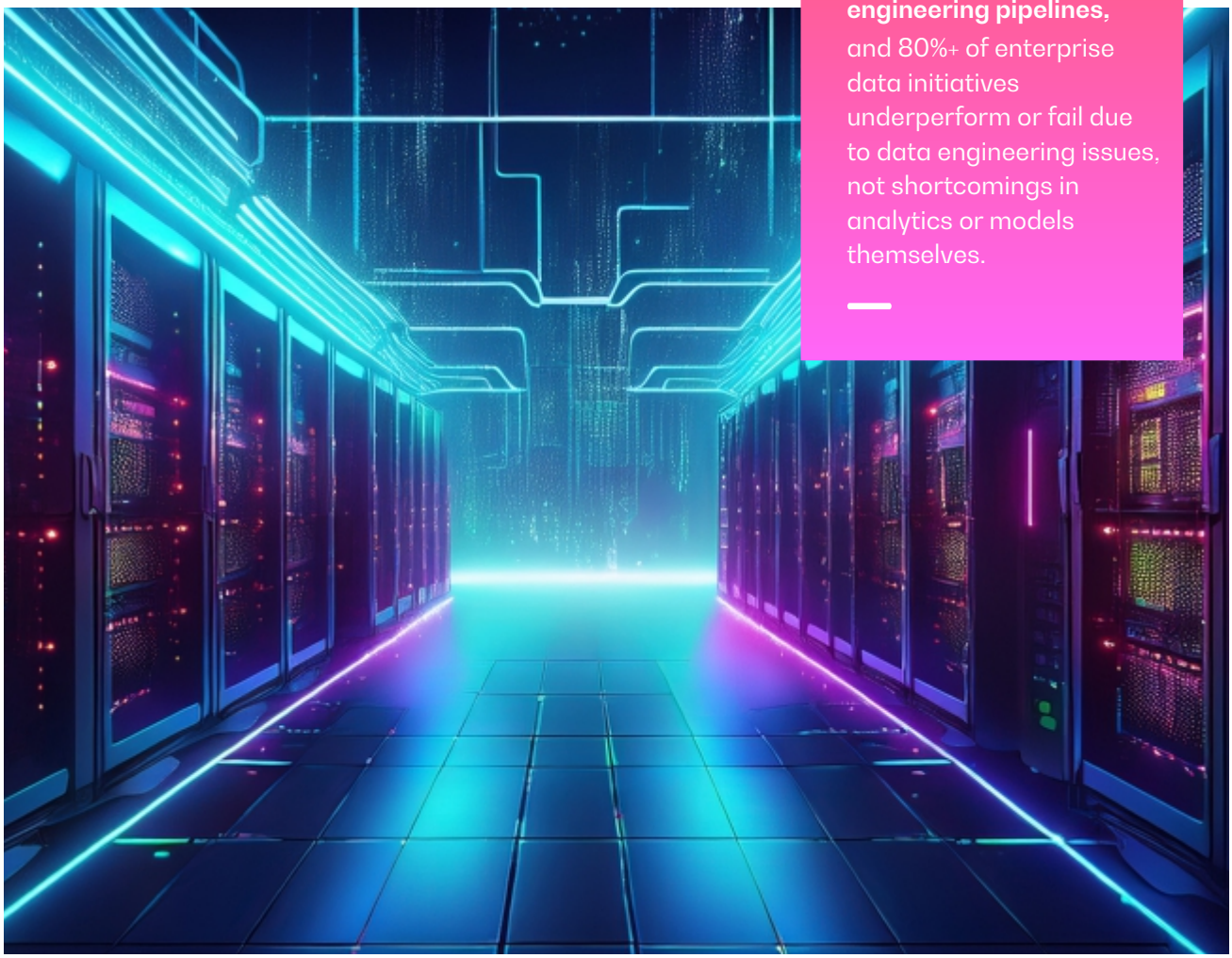
As enterprises race to become AI-first, the spotlight often falls on large language models, generative AI tools, and advanced analytics. Yet, beneath every successful AI initiative lies a less glamorous but far more critical foundation: data engineering. Without a strong data foundation, even the most sophisticated AI models fail to deliver real business value. The right data is a sine qua non for any AI initiative or model.

In fact, a recent study found that

90%

of AI and ML projects

depend directly on data engineering pipelines, and 80%+ of enterprise data initiatives underperform or fail due to data engineering issues, not shortcomings in analytics or models themselves.



An AI-first enterprise is not defined by the number of models it deploys, but by how effectively it can collect, process, govern, and activate data at scale. This is where data engineering becomes the backbone of AI transformation. Through this article, we will uncover how data engineering powers AI at scale and why it has become the true backbone of successful AI-driven enterprises.

01. AI Is Only as Good as the Data Behind It

AI systems learn patterns from data. If the data is incomplete, inconsistent, biased, or outdated, AI outputs will reflect those weaknesses.

According to Gartner, poor data quality costs organizations an average of

$12.9 million per year

in operational inefficiencies and missed opportunities (Gartner).

A common mistake in many analytics and data science efforts is developing models using limited historical data from a small set of stores, relying on manually extracted CSV files, performing one-time data cleaning, and training models in isolated environments.

At first, the results look strong, with high experimental accuracy, polished dashboards, and convincing demos that suggest the AI is working. But when these models move into real operations, the gaps in data, processes, and environments begin to surface, and performance starts to degrade.

To eliminate these issues, data engineering ensures that raw data from multiple sources, like ERP systems, customer interactions, IoT devices, logs, and third-party platforms, is transformed into clean, reliable, and analytics-ready datasets. This process enables AI models to make accurate predictions, generate meaningful insights, and support confident decision-making.

02. The Relationship Between Data Quality and AI Outcomes

AI systems fundamentally depend on the accuracy, completeness, and consistency of data. Models learn patterns from historical and real-time data, and any gaps, errors, or biases in that data directly affect the quality of AI outputs. Even advanced algorithms cannot compensate for unreliable inputs, making data quality a critical determinant of AI success.

Poor data quality has direct and measurable business consequences. This leads to inaccurate predictions, flawed insights, operational inefficiencies, and a reduction in confidence in AI-driven decisions. As trust erodes, many organizations begin to question or even abandon AI initiatives due to unreliable outcomes.

Manual data preparation does not scale and often introduces persistent issues such as duplicate records, missing or inconsistent timestamps, and mismatched product or customer identifiers. When data quality deteriorates, AI outputs become inherently unreliable, aptly captured by the principle of "garbage in, garbage out."

Data engineers, through data cleansing, standardization, validation, and enrichment, ensure that data from multiple sources is consistent and current. By establishing reliable data pipelines and quality checks, data engineering creates a trustworthy foundation that allows AI systems to deliver accurate, scalable, and business-relevant results.

03. Data Engineering Enables Scalable AI

The true value of AI emerges only when it can be operationalized at scale. Scalable AI relies on reliable, real-time data pipelines that can continuously ingest, process, and deliver data across the organization. Modern AI use cases such as fraud detection, personalized recommendations, dynamic pricing, and predictive maintenance cannot depend on delayed or batch-only data.

Data engineering enables these use cases by designing resilient pipelines that can handle high data volumes, multiple data sources, and low-latency requirements.

For example, in e-commerce, real-time pipelines allow recommendation systems to respond instantly to customer clicks, purchases, and inventory changes, directly impacting conversion rates.

According to McKinsey, organizations that successfully operationalize AI through integrated data and analytics workflows are

2.3 times

more likely to outperform peers in terms of revenue growth. However, many AI programs stall because data pipelines are fragile, manual, or disconnected from operational systems.

Beyond real-time processing, data engineering is critical for moving AI from pilots to production. Pilot projects often work with small datasets, manual data preparation, and isolated environments. Production AI, however, requires automated pipelines, standardized data models, data quality checks, and continuous monitoring. Data engineering ensures consistency between training and inference data, reducing errors and improving model reliability. In financial services, fraud detection models only deliver value at scale when transaction data flows securely and continuously into AI systems that can act in milliseconds.

By operationalizing data pipelines and ensuring scalability, reliability, and consistency, data engineering transforms AI from promising experiments into enterprise-grade, production-ready systems that deliver sustained business impact.

04. Data Engineering at Enterprise Scale

A large omni-channel enterprise operating across 24 countries faced high costs, manual processes, and limited transparency in its legacy analytics platform. We migrated the organization to Databricks on Microsoft Azure, redesigning data pipelines using Azure Data Lake and the Medallion architecture to improve scalability, automation, and performance. The result was significantly faster data processing, reduced redundancy, and quicker access to insights for business users highlighting how strong data engineering is critical for enterprises to deliver reliable, scalable, and high-impact analytics.

To know more in detail, **click here**.

05. Why Data Engineering Is Non-Negotiable for AI-First Enterprises

In an AI-first world, a strong data engineering system is non-negotiable. Even the most advanced AI models cannot deliver value without reliable data pipelines, quality data, and scalable data processes behind them. Data engineering is what makes AI accurate, scalable, and production-ready.

06. Conclusion

For enterprise leaders, the key takeaway is clear: becoming AI-first is less about adopting more models and more about building strong data foundations. Organizations that invest in enhancing their data engineering processes and systems are the ones that successfully move AI from experimentation to real, sustained business impact.

07. About the Author

Anirban Majumder

Anirban is a seasoned IT professional with over 20 years of experience working closely with enterprises to design and deliver solutions that address real business needs. With a strong focus on Marketing and Business Development, he partners with leadership teams to align technology initiatives with growth strategies and market outcomes. He is deeply committed to helping organizations unlock tangible value from their data, enabling them to scale analytics and AI initiatives effectively.

About Movate

Movate is a digital technology and customer experience services company committed to disrupting the industry with boundless agility, human-centered innovation, and a relentless focus on driving client outcomes. Recognized as one of the most awarded and analyst-accredited companies in its revenue range, Movate helps ambitious, growth-oriented companies across industries stay ahead of the curve by leveraging its world-class talent of over 12,000+ full-time Movators across 21 global locations and a gig network of thousands of technology experts across 60 countries, speaking over 100 languages.

For more details, please mail us at info@movate.com

www.movate.com

movate®