SPONSORED CONTENT

3 Critical Criteria for AI Success

A fine-tuned focus on data, an embrace of open standards, and the simplification of processes all accelerate the time to value from AI initiatives.

Executive Summary

Companies worldwide are eager to tap the potential of artificial intelligence (AI) and machine learning (ML) to increase innovation and efficiency. These goals can come to life by enhancing customer experiences, accelerating life-saving drug discovery, improving supply chain logistics, as well as other long-term endeavors around space travel and climate change.

However, few organizations report they've been successful at using AI to deliver business value[1]. This is a massive lost opportunity. Time is wasted when businesses fail to move beyond the reactive analysis of "what happened" to preparing a range of predictive options for "what might happen."

This content explores:

- **Three key stumbling blocks** that typically cause many AI initiatives to fail despite the scale and variety of solutions
- **The critical capabilities** that are necessary for success
- **The ways in which the Databricks Lakehouse Platform,** with its data-centric AI approach, helps deliver these capabilities and accelerates time to business value

1 Why AI investments fail to deliver

CIO

databricks

It's crucial that engineers have the ability to observe input data and features in real time.

3 AI Stumbling Blocks

The three major obstacles to AI success are anchored on data, people, platforms and processes.

Obstacle 1
High-friction data management

Before data science, there is data prep, which in itself is a huge challenge. ML models are paramount for accurate predictions, and the quality of training data — its freshness, consistency and transformations — is critical.

Unfortunately, it's difficult to operationalize data for ML today. For example, customer data is often sprawled across silos both on-premises and in the cloud. Operationalizing data pipelines for ML across disparate sources — such as data warehouses and data lakes — is difficult. Over decades, businesses have added significant volumes of data to these resources to meet immediate needs to better understand their customers, optimize supply chains and more. Although these sources may still meet today's requirements, they limit data lineage and visibility, and create barriers to model versioning, governance and compliance.

Another challenge is "featurization." Being able to engineer features in ML models is essential to prediction accuracy, especially to prevent skew. However, data scientists today have to either port all their data over to a siloed storage resource of features or build one themselves. This can result in delays and the potential for stale data by the time it's used in model training.

Also, once an ML model is in production, there is the possibility of drift. This issue typically arises from changes in data or refinement of the predictor element. For example, a model trained on data that captured customer behavior pre-pandemic would be less accurate in predicting behavior during the pandemic. Thus, it's crucial that engineers have the ability to observe input data and features in real time, and be able to seamlessly share model metrics and logs. This can only be done on a platform that provides end-to-end visibility into the ML lifecycle.

Obstacle 2
Pockets of AI applications

While data has proliferated across the enterprise, the power to leverage it has not been spread evenly. Data science and AI/ML are still code-heavy, serve primarily niche use cases, and are out of reach for most analysts and managers in an organization.

These issues can be tied back to two primary challenges: people and tools. Businesses often lack the skilled staff and frameworks to consistently curate, catalog and utilize various types of data in multiple silos and formats across the data lifecycle. This problem is exacerbated by the dearth of data scientists and AI/ML practitioners.

In addition, most data management tools are not designed to be used in a distributed environment that includes in-house data storage resources ("stores") and multiple cloud providers. This makes it more difficult to manage the financial, transactional and management costs of storing data, moving it off disk into memory, and running compute processes against it. In addition, there is no standard tooling for AI/ML.

Obstacle 3
Complex ML operations

ML in production is a long and convoluted process. Just as a plant manager needs a clear, accurate inventory of the capital assets in a factory for efficient use, businesses need to track ML models that turn their data into predictive insights and products.

The hardest problems in the ML lifecycle are data problems. This lifecycle involves multiple teams — including development engineering, data scientists and DevOps teams — each using their own subset or version of data. Consequently, reproducing models at each stage becomes a challenge. For example, it takes an average of 9 months for AI projects to go from pilot to production. Training, tracking, logging, reproducing and sharing hundreds of models over an extended development period is difficult, but without a common underlying data platform, it's almost impossible.

The delay to realizing business value is also due to the complexity of setting up ML-optimized infrastructure, importing packages and libraries for distributed training, and model management. These challenges are not just limited to on-premises architectures. Even in cloud platforms, a considerable amount of compute setup tends to be manual in nature.

For model management, data science and ML teams work on hundreds of models at once. This makes it challenging to track and log thousands of metrics, register and promote the right models, reproduce models at various stages of the lifecycle, share results across teams, and deploy models. Today, ML engineers typically grapple with multiple proprietary tools and workflows for standard ML operations (MLOps).

Also, successful ML projects require massive cross-functional collaboration among line-of-business units, central data engineering and IT teams, and data science teams. When each of these teams use bespoke tools across the ML lifecycle — from data processing and modeling to building deployment and production tools — it becomes challenging to provide a common framework for collaboration. As a result, ML projects get delayed or scrapped.

For model management, data science and ML teams work on hundreds of models at once.

A strong data foundation is critical to achieve a state of predictability.

3 Requirements for Highly Effective AI

Organizations that derive the most business value from their AI investments typically share three characteristics.

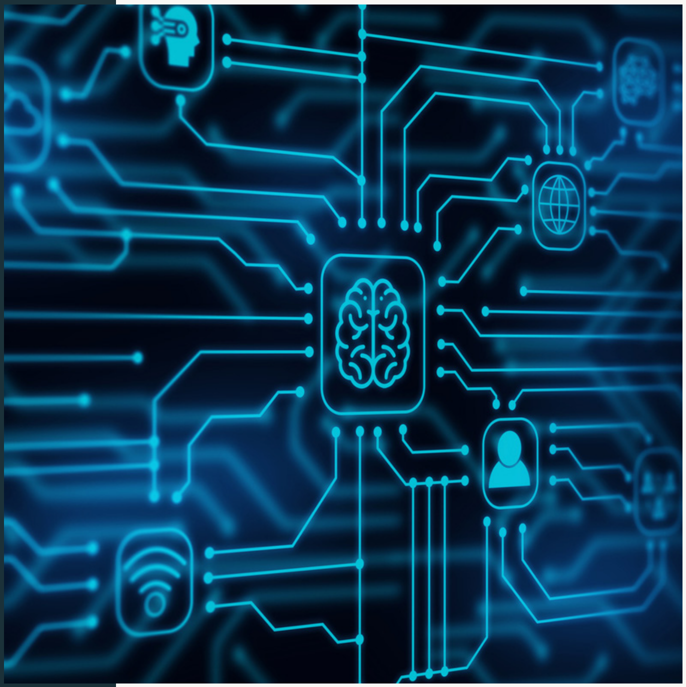
Factor 1
A strong data foundation

AI-driven organizations are moving beyond a reactive “what happened?” analysis of past actions and conditions toward generating high-confidence predictions of what might happen, given various scenarios or options. Using AI to predict the outcomes of possible actions and to assign levels of confidence to those predictions, businesses can better anticipate customer and market needs and respond to them more effectively than their competitors.

A strong data foundation is critical to achieve a state of predictability. This means building a unified, consistent data layer using a platform that supports all data types for AI development. It supports a seamless AI journey and allows organizations to track any issues directly to the data-system source. By bringing together data, models and MLOps on the same platform, organizations become more AI-driven.

A unified platform that evolves and scales for AI allows for the following capabilities:

- **Ensures** that all data can be supported in one place and that it is consistent and traceable
- **Applies** concurrency, consistency and lineage to disparate data types
- **Helps** data science professionals build, transform and visualize features and their interactions prior to downstream MLOps, using the freshest and most reliable data from any source
- **Enables** ML engineers with real-time observability, creates automated and custom alerts, and instantly shares relevant information with DevOps and data science teams
- **Runs** ML workloads that seamlessly ingest and merge data from data warehouses, data lakes, batch and streaming sources across a common security and governance framework
- **Creates** version control across sources — including training, test and real-world data — guarantees transactions, validates schema, and ensures faster inference and superior model performance
- **Enhances** capabilities for feature sharing and discovery among data teams, and ensures that the same feature-computation code and data sources are used for model training and inference
- Continuously **monitors** and **improves** the performance of in-production ML models with automated model retraining to manage drift. ML engineers can define custom KPIs relevant to the business to track model performance and set up metrics to retrain and promote the best models to production.



Proactive AI analysis allows businesses to cost-effectively model and compare possible responses even to unlikely events.

At the highest level, these capabilities enable organizations to proactively plan — which is critical in an era of hard-to-predict events such an economic recession, a global health crisis or geopolitical threat. Proactive AI analysis allows businesses to cost-effectively model and compare possible responses even to unlikely events, helping them cope more effectively with unexpected risks.

For example, global energy giant Shell is using AI/ML not only to meet current needs such as better management of spare parts, but also to develop new energy solutions such as optimizing the layout of wind farms and new ways to charge electric vehicles. Biotech leader Amgen is using a data lakehouse platform to manage its current supply chain, as well as to increase the likelihood of success for future clinical trials.

Factor 2
Scale ML across the organization

To democratize ML, businesses need to consider their approaches to technology and people.

In terms of technology, rarely does any business have all the data required to respond to and predict changing market needs. A single tech vendor cannot offer all the necessary tools to cost-effectively manage and analyze that data. That's why businesses should embrace open standards to store, share, access and analyze data, allowing them to make the greatest use of best-of-breed solutions from multiple vendors. Open standards-based tools also expand the pool of professionals trained in their use and can reduce the time, effort and cost required to gather the data needed to train AI/ML models.

In terms of people, organizations must consider ways to activate the "citizen data scientist." Data consumers should be enabled to easily perform data exploration, create visualizations and generate production-grade ML models for a wide variety of problems without needing to write a single line of code. This allows individuals from across the organization to become data-proficient and leverage the power of data to better inform their team's business decisions.

Factor 3
Simplify MLOps

As organizations become AI-ready, they ultimately need to unify their data and ML models, and provide end-to-end processing, feature creation and model orchestration from experimentation to production. This cycle needs to be repeatable at scale from one to thousands of production models. The lineage is of equal importance. As ML efforts scale, there also needs to be full lineage back to the data-system sources to continuously monitor, tune and improve performance.

The freshest and most reliable data is always available for data science and ML purposes.

All of this requires a simplified MLOps approach, which includes these benefits:

- **ML is data-centric.** ML is an inherent capability of the data platform. There is no need to port data out of a data lake or warehouse into a bespoke ML tool. The freshest and most reliable data is always available for data science and ML purposes, and data engineering and ML teams can operate on the same sharing and governance frameworks. The impact on data teams can be substantial including: self-service data operations for the ML team; fewer back-and-forth processes and delays between data engineering and data science; and significant improvements to productivity and accelerated project timelines.
- **ML can be delivered as a service.** The MLOps approach automates the creation of an ML compute cluster, fine-tuned to the user's preferences. Most popular ML libraries should be included, compatibility-checked and ready-to-go for deployment. Engineers and developers can immediately get started on their integrated development environments or notebooks of choice within the platform.
- **The ML lifecycle can be accelerated.** With a comprehensive and easy-to-use suite of model management capabilities, data teams can quickly explore and wrangle their data in programming languages such as Python, R or SQL; package, track and compare thousands of models; and easily share reproducible versions. In addition, they can seamlessly annotate and transition models across their lifecycle — from development to staging to production. MLOps also automates the creation of optimized clusters with the most popular ML libraries, such as TensorFlow, PyTorch, Keras and XGBoost, as well as libraries required for distributed training such as Horovod. Lastly, built-in visualizations and ad hoc dashboards allow data scientists to quickly visualize and understand the shape and form of the data, and provide error-free transformations, tracking of experiments, and modeling of artifacts.
- **Collaboration between data teams and AI teams can be nurtured.** Simplified MLOps brings together teams and personas across data engineering, data science, production ML, DevOps, and data analysis on a common operational framework for any workload. Using the same data platform, they can interact with comprehensive version tracking and shareability.

Delta Lake provides faster, less expensive delivery of business insights.

The Way Forward: Databricks Prepares Organizations for AI Success

The Databricks Lakehouse Platform helps businesses meet these three criteria for AI success. It simplifies and speeds AI while enabling the collaboration required to move from a reactive to a proactive response. Its data-centric approach combines the reliability, performance and governance of data warehouses with the scalability of data lakes for flexible, high-performance analytics, data science and machine learning.

The Databricks Lakehouse Platform provides:

- **Simplicity:** The Delta Lake open-format storage layer, on which the Databricks Lakehouse Platform is based, simplifies data architectures by replacing silos with one cloud-based platform. It eases integration and portability through the support of open-source solutions such as MLflow, Apache Spark™ and open data standards, while enabling fine-grained access controls for data governance. This reduces costs and speeds time to value for analytic insights.
- **Cost reduction:** Delta Lake dynamically changes the size of data partitions — how data is distributed and accessed among servers — for the best combination of cost and performance. It also continually finds the most cost-effective cloud provider for every data management process, and provisions and de-provisions clusters as needed to minimize costs. As a result, Delta Lake provides faster, less expensive delivery of business insights, along with reduced costs for computing, storage and the retrieval of data needed for AI applications.
- **End-to-end ML lifecycle management:** Databricks provides data science teams with a complete and comprehensive solution to reduce time to deployment with its fully managed and hosted version of MLflow, an open source ML lifecycle management platform. It compresses the ML lifecycle by months, providing an end-to-end platform for data scientists and ML engineers to build, productionize and monitor projects. Databricks allows data science teams to collaboratively track experiments, record and compare parameters and results, and package and deploy models from all popular ML libraries to any serving endpoint. Data science teams can collaboratively build, share and transition models across various stages of the model lifecycle with versions and annotations, sync notebooks to the organization's preferred Git repository, and trigger alerts as needed through the model lifecycle.

Support for open source standards makes it easy for developers and data scientists to share data and jointly develop AI applications.

- **Collaboration:** Support for open-source standards makes it easy for developers and data scientists to share data and jointly develop AI applications. With collaborative notebooks, data engineering, data science and ML, teams can leverage support for multiple programming languages, built-in data visualizations and automatic versioning. In addition, the MLflow Model Registry is a collaborative hub where teams can share thousands of ML models during experimentation, testing and production. It expedites the sharing of expertise and knowledge through a central tracking registry.
- **Governance:** The Databricks Unity Catalog gives users visibility into all the data assets owned by an organization, whether in a data lake, data warehouse or other sources. Organizations can centrally share, audit, secure and manage structured and unstructured data such as tables, files, models and dashboards. In addition, Unity Catalog works in concert with existing data, storage and catalogs, so organizations can build on what they already have and benefit from a future-proof governance model.
- **Data consistency and accuracy:** Delta Lake's strict schema validation helps administrators understand the data they are accessing and its lineage to assure accuracy. The Delta Lake foundation uses transaction logs to ensure that every operation, whether streaming or batch, fully succeeds or is halted and reported to prevent future problems. Its "time travel" capability gives data administrators the ability to review past transactions to determine which ones executed successfully.
- **Performance:** The Delta Lake Delta Engine leverages indexing to provide high performance for business intelligence, reporting, data science and ML from a single source of data. Databricks SQL, built on top of the Lakehouse architecture, is fast and provides low-latency performance.

Become AI-driven

Regardless of the global economic, geopolitical, or health climate, AI applications are helping organizations streamline operations, speed product development, improve the customer experience and more.

To become more AI-driven and ML-ready, here are links to further resources:

- The Future of the Modern Data Stack: The CEOs of Databricks, dbt Labs and Fivetran share their vision
- 5 Steps to a Successful Data Lakehouse by Bill Inmon, the father of the data warehouse
- Additional resources at the Databricks website

Rolls-Royce's platform has helped it extend the time between maintenance for some engines by up to 50%.

Rolls-Royce turns to digital twins to improve jet engine efficiency

By Thor Olavsrud, Senior Writer, CIO | June 10, 2021

Say the name Rolls-Royce and most people think of automobiles, but the British multinational aerospace and defense company has been out of the car business since Rolls-Royce Motors was sold off in the 1970s. Today, Rolls-Royce Holdings is the second-largest maker of aircraft engines in the world, with a foot in marine propulsion and energy as well. Its engines are used in fighter jets, business jets, and more than 50% of long-haul planes.

Now the company is deploying digital twin technology, analytics, and machine learning to dramatically reduce the amount of carbon its aircraft engines produce while also optimizing maintenance to help its customers keep their planes in the air longer.

"Rolls-Royce has been monitoring engines and charging per hour for at least 20 years," says Stuart Hughes, chief information and digital officer at Rolls-Royce. "That part of the business isn't new. But as we've evolved, we've begun to treat the engine as a singular engine. It's much more about the personalization of that engine."

Using its Intelligent Engine platform, the company monitors how each engine flies, the conditions in which it's flying, and how the pilot uses it.

"We're tailoring our maintenance regimes to make sure that we're optimizing for the life an engine has, not the life the manual says it should have," Hughes says. "It's truly variable service looking at each engine as an individual engine."

Rolls-Royce's platform has helped it extend the time between maintenance for some engines by up to 50%, thereby enabling it to dramatically reduce its inventory of parts and spares. Perhaps most importantly, however, it has greatly improved the efficiency of its engines, saving 22 million tons of carbon to date, according to the company. Rolls-Royce is even using AI to better understand how to handle metal scrap and waste from parts when they reach the end of their lifespan.

Product Engineering: Amplified on the Cloud

"Since 2014, we've helped one of our airlines avoid 85 million kilograms of fuel and over 200 million kilograms of carbon dioxide," Hughes says. "We did that by taking data on how the pilot is flying the plane, how the plane is operated, how they do the operational funding around that. We found data and insights that helped them to make better decisions. In areas where they felt there were barriers to change, we helped them design new policies, new procedures."

Under the hood

To fuel its Intelligent Engine platform, Rolls-Royce is using a combination of two-way, real-time data captured from its engines as they fly, and larger datasets captured in batch after planes land, to power its analytics. It feeds the data into a Microsoft Azure data lake and then into a Databricks Lakehouse, where it can be used with Databricks machine learning and AI tools. (Databricks uses the term "lakehouse" to refer to its open architecture that combines the features of a data lake and data warehouse.)

Between the real-time data and the data collected after landing, each flight generates about half a gigabyte of data. The real-time data is used for the company's "Engine Condition Monitoring" service, which analyzes the data for irregularities in engine performance for the purpose of predictive maintenance. The analytics can determine in-flight whether a full inspection will be necessary upon landing, helping the airline plan ahead and minimize travel disruptions. The other data can be used for more detailed predictive modeling.

"We're using that data to check that the engine is still within all our quality and safety tolerances, but also to understand how the pilot has flown that engine," Hughes says. "That means we can offer to extend the maintenance window on that engine longer for a specific customer. The benefit to the customer is the customer sees less interruption because the engine is on the plane for longer, so they can use it more. The benefit for us is that we can optimize how we actually do the maintenance."

Click here for the full article.

Machine learning systems are prone to latencies if the underlying architecture lacks an operational approach to ML.

Best practices to embrace an 'MLOps' mindset

By Yash Mehta, Contributor, CIO | June 28, 2021

Moving an AI project from ideation to realization is a vicious loop, and there is only one way to resolve it – don't let the loop begin! That is true because data deserves expert handling at all levels. Starting with extracting it from different sources to cleaning, analyzing, and populating it, machine learning systems are prone to latencies if the underlying architecture lacks an operational approach to ML – known as MLOps.

Most AI projects do not make it to production due to a gap that sounds very basic but has a massive impact: improper communication between the data scientists and the business. In trying to bridge this gap organizations are looking to MLOps.

MLOps best practices focus on:

- **Providing end-to-end visibility** of data extraction, model creation, deployment, and monitoring for faster processing.
- **Faster auditing and replicating** of production models by storing all related artifacts such as versioning data and metadata.
- **Effortless retraining** of a model as per varying environment and requirements.
- **Faster, securer, and more accurate testing** of the ML systems.

However, developing, implementing, or training ML models was never the main bottleneck. Building an integrated AI system for continuous operations in the production environment, without any major disconnects, is the actual challenge. For example, organizations that have to deploy ML solutions on demand have no choice but to iteratively rewrite the experimental code. The approach is ambiguous and may or may not end in success.

That is exactly what MLOps tries to resolve.

Put simply, DataOps for ML models is MLOps. It is the process of operationalizing ML models through collaboration with data scientists to achieve speed and robustness. A company called Neuromation has a complete service model wrapped around strategizing the MLOps. The ML services provider emphasizes bringing data scientists and engineers together to attain a robust ML lifecycle management.

Apart from data scientists, the collaboration includes engineers, cloud architects, and continuous feedback from all stakeholders. Along the way, it emphasizes implementing better ML models in the production environment and creates a data-driven DevOps practice.

Click here for the full article.