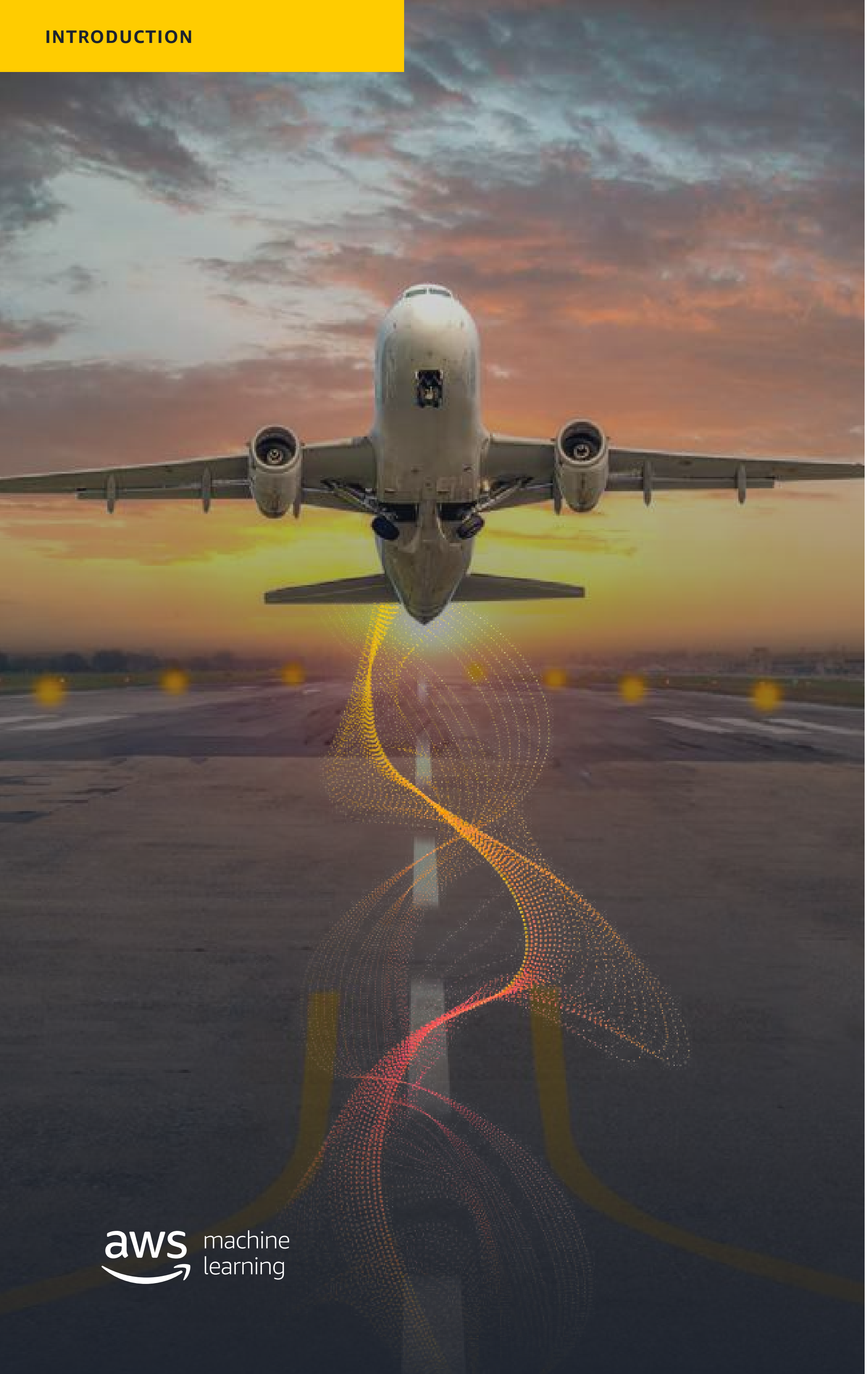
aws machine learning

The Machine Learning Journey

The path toward leveraging the full power of machine learning technologies



Forging ahead

When deployed with the right strategies, machine learning (ML) can increase agility, streamline processes, boost revenue by creating new products and improving existing ones, and enable better, faster decision making.

There's no doubt machine learning and artificial intelligence (AI) can help companies achieve more—in a recent survey by McKinsey, 63 percent of respondents reported revenue increases from AI adoption in the business units where their companies use AI.[1] It's also clear that adopters continue to have confidence in AI technologies' ability to drive value and advantage. According to Gartner, 75 percent of enterprises will shift from piloting AI to operationalizing AI by the end of 2024,[2] and according to Deloitte, 57 percent say AI will transform their organization over the next three years.[3]

While machine learning has been around for decades, its accessibility as a tool to transform businesses is relatively new. And the lack of a singular proven path to machine learning success is keeping some businesses waiting on the sidelines, unsure of how to take the next (or even the first) step on the journey. This eBook is designed to help businesses forge ahead, outlining a proven path from the first step to measuring results—with insights from Amazon's own machine learning heritage and its experience helping thousands of customers realize their own initiatives.

It's time for organizations to overcome their machine learning worries, stop playing catch-up, and forge ahead with confidence. No matter where organizations are on their machine learning paths, they'll find the guidance they need to take the next step to machine learning success.

[1] https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact

[2] https://www.gartner.com/en/newsroom/press-releases/2020-06-22-gartner-identifies-top-10-data-and-analytics-technolo

[3] https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/state-of-ai-and-intelligent-automation-in-business-survey.html

What are artificial intelligence and machine learning?

You've probably heard artificial intelligence (AI) and machine learning described in a number of ways, so let's take a step back and review their exact definitions:

Artificial intelligence (AI) is a way to describe any system that can replicate tasks that previously required human intelligence.

Almost always, this is related to some kind of complex decision making task where human judgment would normally be required. Most use cases for AI are looking for a probabilistic outcome—making predictions, classifications, or decisions with a high degree of certainty and in ways that are similar to human judgment.

Almost all AI systems today are created using machine learning. **Machine learning** uses large amounts of data to create and validate decision logic. This is known as a model. The AI system feeds input data into that model, and then the model outputs human-like predictions or classifications. Essentially, machine learning is **the underlying technology that powers intelligent systems**.

AI can be created without machine learning, but right now, machine learning is the primary method for creating AI systems. Similarly, machine learning can be used for more than AI, but right now, the majority of machine learning is AI-related.

Why machine learning?

Before digging into the steps of the machine learning journey, let's explore why businesses should go on that journey in the first place. After all, even with the guidance in this eBook, completing the steps outlined here will require continued investments and unwavering dedication. Businesses will need to regularly remind themselves what they're fighting for—keeping their eyes on the precise business benefits that can be unlocked by fully leveraging machine learning technology.

Businesses already realize the impact of:

1

Optimizing business with new efficiencies

Machine learning can be used to create greater efficiency through sophisticated demand planning and forecasting models. While this is true in almost every industry, retail provides some specific evidence. AI-based forecasting is reducing lost sales due to product unavailability by up to 65 percent and resulting in two million fewer product returns per year.[4] Using an AWS-based predictive ordering solution, **Domino's Pizza Enterprises Limited** is delivering on an initiative to have pizza ready for pickup within 3 minutes of ordering or safely delivered within 10.

2

Smarter, faster decision making

Informed by data and analytics sources that grow smarter through machine learning, businesses and their workforces can make more informed, faster decisions that allow them to act on opportunities sooner and get better results. **T-Mobile** customer service agents use AI to quickly access the information most salient to customer needs. By providing agents with contextual information in real time, T-Mobile helps guarantee that each customer's issues are quickly and accurately resolved.

[4] https://www.mckinsey.com/~/media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx

aws machine learning

3 Adding new capabilities to existing products

Machine learning can enrich existing products, improving customer engagement and attracting new users through deeper experiences. For instance, **Livongo** is a platform and mobile app that works with smart devices, such as a connected blood glucose monitor, to help people manage medical issues. It uses machine learning to translate data from blood glucose readings, physical activity, and meal logs, in addition to smartphone data and other important data, into timely and actionable "health nudges." These personalized messages around diet, exercise, medications, and more—delivered in real time to members on their connected devices—help them avoid complications that could land them in the hospital, saving the system (and themselves) money.

4 Inventing net-new products

With machine learning, businesses can leverage data to develop revolutionary ideas and bring new products (and even new categories) to market. **Convoy** disrupted the trucking industry by introducing a machine learning-powered model to automate logistics. Convoy's solution provides better matches for shippers and truckers, allowing them to move freight more efficiently—and lowering costs for both parties.

Now that we've outlined the "why" of machine learning, it's time to explore the "how." The next sections will demonstrate the steps of the machine learning journey, using Amazon's own path and those of AWS customers to exemplify the necessary changes that must take place in order to successfully implement, deploy, and scale machine learning.

The machine learning journey

The machine learning journey is not always a straightforward path. Achieving success with machine learning requires more than just great technology—it also means ensuring the organization is aligned to the right goals. Identifying and reaching those goals will necessitate broad changes in processes, management, and culture. The next sections will explore how organizations can overcome common challenges that often impede progress and take the right steps to implement machine learning in efficient, sustainable ways.

1 Championing a machine learning culture

According to Gartner, the global business value derived from artificial intelligence (AI) is projected to reach $3.9 trillion in 2022.[5]

"(Machine learning) will empower and improve every business, every government organization, every philanthropy—basically there's no institution in the world that cannot be improved with machine learning," said Jeff Bezos, founder and CEO of Amazon.

Unlocking the full business potential of machine learning will require cultural changes in team organization, objectives, and outlook.

In order for machine learning to proliferate through the organization, both business and technical teams must work together and share the same priorities. To achieve this, at the outset, the machine learning effort must be supported from the highest levels, with goals set by executive champions and an investment in the technology and the processes that enable success.

It's important that management take a wide-scale view while fostering machine learning initiatives. Executives must be firm in their goals but flexible in how the organization reaches them. Mistakes are sure to be made. By staying focused on the long-term outlook and not allowing discouragement, organizations can glean wisdom from every error and apply those learnings to champion a machine learning culture throughout the business.

Perhaps the largest cultural change organizations must undergo is becoming fault-tolerant. Machine learning is an iterative process, one which can only succeed through constant experimentation. More often than not, these experiments will result in failure. Only by tolerating these failures—and refusing to grind progress to a halt in the name of determining "what went wrong?"—can organizations consistently reach the breakthrough successes waiting on the other side.

[5] https://www.gartner.com/en/newsroom/press-releases/2018-04-25-gartner-says-global-artificial-intelligence-business-value-to-reach-1-point-2-trillion-in-2018

How Amazon did it

Amazon has been using machine learning in the business for over 20 years. But more than 10 years ago, to further adoption of machine learning across the company, Amazon's leadership team asked every business leader in the organization—irrespective of whether they ran a research team, a fulfillment center, or an HR department—to answer the question of how they planned to leverage machine learning in their businesses.

In most cases, "We don't plan to" was not an acceptable answer. This forced leadership, domain experts, and technical experts to collaborate on machine learning initiatives and to let nothing halt their progress—even in instances where tangible benefits were still years down the road.

In addition to hiring data scientists externally, Amazon also created the Machine Learning University, which trained many of its developers to use machine learning more effectively. The company built tools like Amazon SageMaker, which simplifies the process of creating models and lowers the barrier to entry—so machine learning could scale more effectively.

Now, more than10 years later, there isn't a single department at Amazon that hasn't been touched by machine learning. Amazon's personalization technology that provides recommendations to customers has significantly improved since its first model 20 years earlier and has been applied to other areas of the business.

The company uses machine learning throughout its fulfillment process and leverages a forecast system that can predict demand for nearly every product in its enormous inventory. These prediction models allow Amazon to better deliver on customer expectations of convenience, cost, and delivery speed.

"We forecast millions of products every single day across all of our Amazon sites worldwide," said Jenny Freshwater, director of forecasting at Amazon. "And without machine learning, we would not be able to produce those forecasts."

Amazon has also developed technology to give consumers an entirely new way to interact with technology through Alexa. The company has developed groundbreaking technology with autonomous flight via Amazon Prime Air drones and uses robotics in its fulfillment centers to get packages to consumers faster. Plus, **Amazon** is using machine learning to minimize the amount of packaging that customers have to dispose of, reducing the weight of their outbound packaging by 33 percent and eliminating 915,000 tons of packaging material worldwide.

Achieving these successes required great investments in technology, research, and talent. But those investments would have gone to waste without the cultural change that pushed them forward through many failures and unexpected challenges. Every organization must foster this same fault-tolerant culture of experimentation and innovation before the machine learning journey can truly begin.

2 Rethinking data strategy

Machine learning successes are highly dependent on quality data. Without a proper data strategy in place, progress will slow to a crawl and hamper the effectiveness of the final model. Worse yet, if the model is informed by bad data, the results it generates may be misleading—or even flat-out wrong.

"(Machine learning models are) highly sensitive to data quality," Freshwater said. "So we learned—in many cases the hard way—that the time spent on getting data of high quality on the way in paid dividends in our models on the way out."

The right data strategy for machine learning should aim to break down silos, enabling IT teams to easily, quickly, and securely access and collect the data they need.

While modern data strategies take many forms, data lakes are becoming an increasingly popular core component of the most efficient models. Data lakes offer more agility and flexibility than traditional data management systems, allowing companies to manage multiple data types from a wide variety of sources and to store the data—whether structured or unstructured—in a centralized repository.

Once stored, the data can be analyzed by many types of analytics and machine learning services—faster and more efficiently than with traditional, siloed approaches. Data lake architectures also enable multiple groups within the organization to benefit from analyzing a consistent pool of data that spans the entire business.

For help developing a more holistic data strategy that includes data lakes, interact with the **AWS Data Flywheel**.

"We are using AWS data-analysis technologies to predict … precisely how fast converting lines should run to avoid tearing. By reducing paper tears, we have increased profits by millions of dollars for one production line."

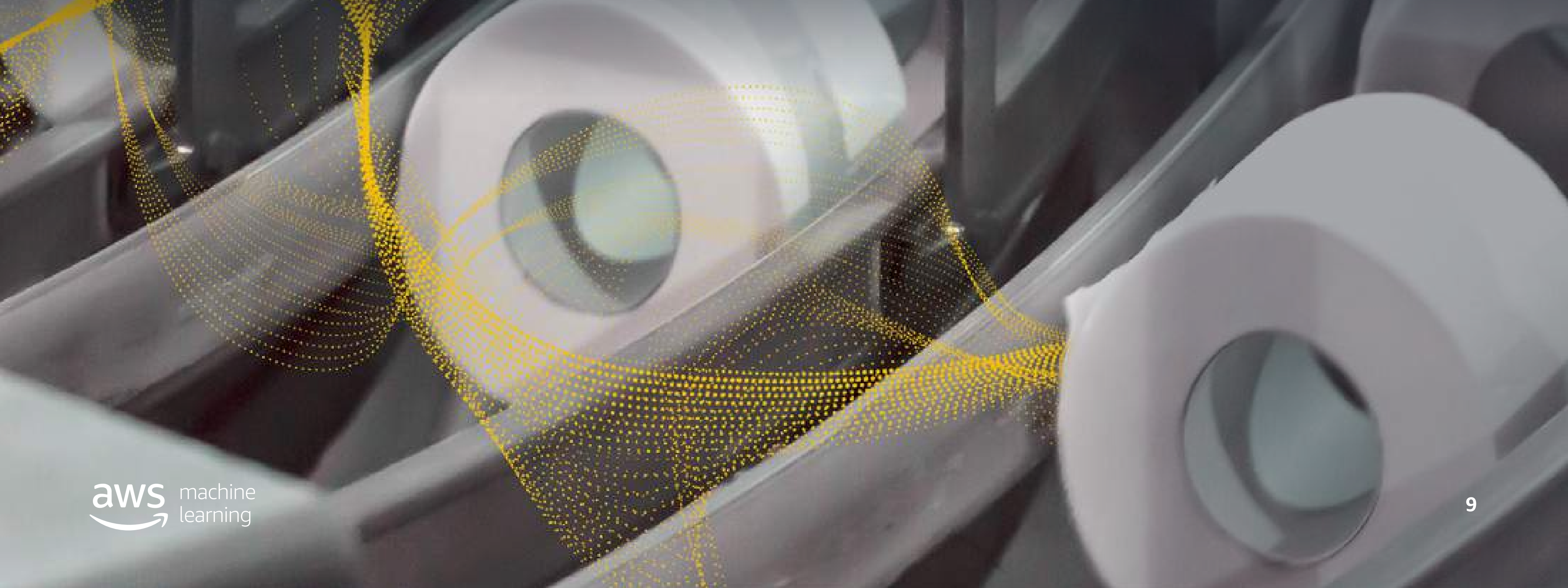
Steve Bakalar, Vice President of IT/Digital Transformation, Georgia-Pacific

How Georgia-Pacific did it

Hundreds of paper and tissue parent rolls are produced every day at Georgia-Pacific manufacturing facilities across North America. If tears or breaks occur frequently, it leads to paper-machine and converting-line downtime that can cost Georgia-Pacific millions of dollars per year per line.

Georgia-Pacific started by migrating 50 TB of structured and unstructured production data from a legacy database infrastructure to a cloud-based data lake. By layering AWS databases and analytics tools on all of that data, Georgia-Pacific was able to optimize key manufacturing processes to predict equipment failure 60–90 days in advance. By reducing paper tears and unplanned downtime, the company was able to increase profits by millions of dollars for one production line.

Read the full story »



3 Finding the right business problem to address

One mistake organizations often make in their machine learning journeys is employing discrete data scientists who work in silos to build machine learning models as proofs of concept—rather than to solve real business problems. With no specific business problem to solve, IT executives will find it increasingly difficult to demonstrate the value of machine learning projects to their business executive counterparts. This can stall or even stop progress on machine learning initiatives.

Here are some important questions organizations should ask before embarking on their machine learning journeys:

1. Is the project important enough to get attention and adoption?
2. Does it solve a real business problem?
3. Are there places where the organization already has a lot of untapped data?
4. Does the project require machine learning?
5. Can it be done by a single business?
6. Can it eventually be operationalized?

“A first step is to identify a problem that is rich in data, but (one that) you haven’t been able to solve through traditional methods,” Freshwater said.

In a successful machine learning journey, organizations create machine learning teams that are built to address specific business problems. This requires including both technical and domain experts within these teams. While the technical experts will take on the brunt of model creation, they need the field knowledge of domain experts to define precise business challenges and identify the data most important to finding a solution.

This approach is also critical to change management—when technical and domain experts collaborate to create machine learning models, employees will feel more confident in making decisions based on the algorithm’s logic.

Together, these teams should also work through how to measure success. “Make sure you . . . have very crisp and clear metrics as you embark on the machine learning journey,” Freshwater said. “Many times, your models are taking over for something existing and you want to make sure that they’re actually better and that you can measure it.”

For more on measuring the success of machine learning initiatives, refer to **Step 6** in this eBook.

Some organizations have the talent in-house to identify the problems that would be best addressed by machine learning and to implement the appropriate pilot programs. Organizations that need help in this area should reach out to experts and collaborate with them to “work backward” from business challenges—and then go step-by-step through the process of creating machine learning projects to solve them.

aws machine learning

How the NFL did it

For decades, the NFL has worked to provide deeper insights into its players and teams, to satisfy both the need for improved player safety and the insatiable fan appetites for data and statistics.

To address this need, the NFL worked with AWS to create the machine learning-powered NFL Next Gen Stats (NGS). Since data science and football are wildly different disciplines, the NFL wisely included both technical and domain experts in the creation of NGS, ensuring both groups could work hand in hand to identify the right data and develop stats.[6]

Leveraging RFID tags to track player movement, NGS provides real-time location data, speed, and acceleration for every player during every play on every inch of the field. By simulating different situations within a game environment, the NFL aims to foster a better understanding of how to treat and rehabilitate injuries in the near term and eventually predict and intervene to prevent injuries in the future.

NGS also uses machine learning models to calculate more than 20 different advanced statistics that are compelling to fans. One example is the Expected Rushing Yards stat, which is designed to show how many rushing yards a ball carrier is expected to gain on a given carry based on the relative location, speed, and direction of blockers and defenders.

Insights like Completion Probability wouldn't exist without the partnership between technical experts—who can build and train the models that crunch the necessary data—and domain experts—who know what data to measure to create the most exciting statistics.

This partnership also helps to build acceptance for NGS, as broadcasters are more likely to cite advanced stats that football experts (and in some cases, the broadcasters themselves) had a hand in creating.

Read the full story »

[6] https://aws.amazon.com/nfl

4 Upskilling your teams

In parallel with creating a data strategy, organizations must focus on arming their engineer teams with the right skills.

Organizations are growing increasingly aware of the IT skills gap—the expanding separation between technologies and the ability of internal IT specialists to take full advantage of them. Closing this gap for machine learning will require a combination of training and recruiting. The reality is, there aren't enough data scientists today to lead the machine learning transformation that is coming. This requires organizations that want to leverage machine learning to first invest in developing their talent.

While there is no one-size-fits-all solution to the machine learning skills gap, there are proven methods that can maximize the abilities of existing staff, reducing the need to make large investments in buying or borrowing pre-trained expert talent. These methods include:

Defining the skills gap: Before closing the skills gap, an organization must identify the precise differences between what it needs or wants its employees to do and what its employees currently have the ability to do.

Understanding how skills are mapped: Since machine learning initiatives are interdisciplinary efforts, an organization should map the skills needed across data scientists, machine learning specialists, application developers, statisticians, and other subject matter experts in the business.

Customizing training for specific needs: If an organization has existing training curriculums that could be useful, it should work to tailor those materials to the business' specific machine learning needs. Leaders should also investigate pre-trained AI services that provide ready-made intelligence for business applications and workflows.

In addition to training, you'll need to align teams to successfully tackle machine learning problems. This includes:

Promoting a culture of empowered teams: Machine learning project teams must be cross-functional, possessing the authority to execute individual objectives and the freedom to organically cross-pollinate with other teams as demands dictate and opportunities arise. To make this kind of teamwork possible, management will need to embrace new structures—letting go of the strictly hierarchical and departmentally siloed organizational models of the past.

Starting with a pilot team: Establish a pilot team of engineers and task it with a machine learning project. "I'd recommend putting a couple of really smart people on trying to figure out what metrics you want to optimize for or predict . . . just start really small," Freshwater said.

Enabling organic transformation: Once the pilot project is complete, the business can split up the team, add new engineers to create new teams, and task them with new projects. This process continues, allowing knowledge to organically spread from veteran team members to new recruits and pollinate between teams.

By following this guidance, many organizations are finding that the people they currently have actually are the people they need to close their machine learning skills gaps. While some recruiting may still be required, organizational, process, and management changes can do much of the work to upskill talent for machine learning success.

How Morningstar did it

Investment research firm Morningstar uses machine learning to automate data collection processes and expand the number of funds it covers. The company does this by leveraging predictions from a machine learning model trained to emulate Morningstar analysts' fund evaluation process.

To train its employees and accelerate machine learning application, Morningstar uses AWS DeepRacer—a tool that facilitates hands-on machine learning training through a fully autonomous 1/18th scale racecar driven by reinforcement learning, a 3D racing simulator, and a global racing league. More than 445 Morningstar employees from multiple functions and eight countries—including 35 percent of its technology function—have been engaged in the DeepRacer League.[7]

Morningstar has dozens of machine learning projects in the pipeline for 2021. These include a reinforcement learning program that searches for patterns in regulatory filings and an algorithm that identifies and fixes broken links to the websites of financial institutions.

"Our DeepRacer challenge harnesses our employees' enthusiasm for machine learning and artificial intelligence. It provides hands-on training across the company and accelerates Morningstar's practical application of machine learning across our investing products, services, and processes. The response from our teams has gone well beyond my expectations, and it has been a fun way to unite our global teams, whether in technology or other functions."

James Rhodes, Chief Technology Officer, Morningstar

[7] https://newsroom.morningstar.com/newsroom/news-archive/press-release-details/2019/Morningstar-Launches-Global-AWS-DeepRacer-Corporate-Competition-to-Accelerate-Application-of-Machine-Learning/default.aspx

5 Scaling beyond pilot projects

After the first few successful pilots, organizations must take the next step on the journey: sustainably scaling machine learning across the business. This is both a technical and a cultural challenge.

Achieving scalability requires organizations to make it easier for their developers to use machine learning. Building machine learning models at scale can be labor intensive and complex, which can slow innovation.

Many organizations are solving scalability with Amazon SageMaker, an end-to-end solution that covers the entire machine learning workflow to build, train, and deploy machine learning models. By using Amazon SageMaker, organizations can get their models into production faster and at a lower cost, enabling sustainable expansion of machine learning initiatives beyond pilot projects.

There are several ways companies approach the cultural shift necessary to scale machine learning. Some might find success by creating a center of excellence that rallies the community and continues to push for new initiatives. Or, like Amazon, organizations can make machine learning an integral part of yearly planning processes, continuously bringing domain and technical experts together to brainstorm and determine the company's next step.

How Intuit did it

Using Amazon SageMaker, Intuit reduced machine learning deployment time by 90 percent—from six months to one week. By centralizing its machine learning initiatives, Intuit fosters innovation and deploys AI and machine learning techniques at speed and scale—achieving business value that goes beyond its products and services.

“AWS gives people within Intuit a common platform to share and collaborate with data in a secure environment,” said Ashok Srivastava, senior vice president and chief data officer at Intuit. “For example, Amazon SageMaker gives us the platform and infrastructure we need to apply our sophisticated AI and machine learning technologies.”

Watch the video »

“AWS gives people within Intuit a common platform to share and collaborate with data in a secure environment.”

Ashok Srivastava, Senior Vice President and Chief Data Officer, Intuit

6 Measuring the results

When measuring the results of machine learning efforts, the traditional "project ROI" viewpoint—where a project has a defined start and end point, a budget, and a return—is reductive and can be detrimental to the initiative's success. If the project doesn't generate a positive return within the given time frame, the business may lose interest and miss out on critical opportunities down the line.

Instead, executives and IT alike must measure machine learning efforts based on what success means for their business with regard to the processes being optimized. In addition, they must view machine learning efforts as long-term investments, acknowledging that a true "return" may not be realized for several years and throughout countless iterations.

When planning machine learning initiatives, it's better to view the process through the lenses of agility, competitive advantage, and/or risk tolerance rather than expected return. Organizations will have greater success if they disregard the question of "What will be my return on investment in X months?" in favor of something more like "If we don't invest in this now, will we fall behind our competitors in X years when the technology matures?"

While traditional ROI metrics may not be the best approach, the business impact of machine learning initiatives can still be measured—it just requires a different outlook.

Machine learning results can be measured through something resembling a "value tree," where the main trunk of the tree represents the traditional "revenue return" and branches extending from the trunk recognize the value of other business outcomes.

The specific branches of the value tree will depend on the organization, the industry, and the initiative, but they might be things like "time saved through automated processes," "new leads, markets, and opportunities identified," "customer service improvements," and/or "increases in upsells."

Measuring the success of machine learning through a more holistic and long-term model will keep your teams focused on the best outcomes for the future of the company.

Taking the next step with AWS

No matter where organizations are in their machine learning journeys, AWS provides products, solutions, and services that can help them take the next step. Featuring the world's broadest and deepest set of machine learning and AI services, AWS has worked with over ten thousand customers to help them successfully implement machine learning.

AWS is dedicated to putting machine learning in the hands of every developer and is working tirelessly to solve the toughest challenges that stand in the way of that goal. AWS capabilities are built on the most comprehensive cloud platform, are optimized for machine learning with high-performance compute, and compromise nothing in security and analytics.

Let's explore current machine learning offerings from AWS—and see how they can help organizations progress in their journeys.

Amazon SageMaker: Amazon SageMaker enables developers and data scientists to quickly and easily build, train, and deploy machine learning models—thus simplifying scalability across the entire business. Amazon SageMaker removes the complexity that gets in the way of successfully implementing machine learning across use cases and industries—from running models for real-time fraud detection to virtually analyzing biological impacts of potential drugs to identifying the best driver in F1.

Machine learning with AWS, by the numbers

AWS machine learning solutions:

- Reduce training time by **50%**[8]
- Provide **90%** scaling efficiency[9]
- Deliver **3x** faster network throughput[10]
- Improve price and performance by **25%**[11]

91% of cloud-based PyTorch runs on AWS

92% of cloud-based TensorFlow runs on AWS

[8] As measured in the ResNet-50 benchmarking test, AWS-optimized TensorFlow recorded the fastest training time, by over 50%

[9] Using AWS-optimized TensorFlow allows for near-linear scaling efficiency, up to 90% compared to 65% using stock TensorFlow

[10] than other providers using P3dn instances

[11] using C5 instances powered by 3.0GHz Intel Xeon compared to previous generation instances

AWS AI Services: No prior machine learning experience is required to take advantage of these AI-powered services from AWS:

Personalization	Advanced text analytics	Voice
Forecasting	Conversational agents	Transcription
Image and video analysis	Translation	Document analysis

ML Frameworks: AWS customers can choose from TensorFlow, PyTorch, Apache MXNet, and other popular frameworks to experiment with and customize machine learning algorithms. They can use the framework of their choice as a managed experience in Amazon SageMaker or use AWS Deep Learning AMIs (Amazon Machine Images), which are fully configured with the latest versions of the most popular deep learning frameworks and tools.

Infrastructure: AWS customers benefit from a broad set of powerful compute options, ranging from GPUs for compute-intensive deep learning to FPGAs for specialized hardware acceleration to high-memory instances for running inference. Amazon EC2 provides a wide selection of instance types optimized to fit machine learning use cases—regardless of whether customers are training models or running inference on trained models.

Learning Tools: AWS also offers a number of learning tools and services to help organizations improve their machine learning capabilities, including:

AWS DeepRacer	Machine Learning Training and Certifications
AWS DeepLens	Amazon Machine Learning Solutions Lab

Solving the biggest machine learning challenges

Most organizations have made some investments in machine learning and are at some stage of the journey. But many find themselves hitting stopgaps along the way, worried that costs and complexities will grow too high as they progress.

In this eBook, we explored the steps toward forging ahead and realizing the full power of machine learning. To recap, let's look at the biggest challenges we identified along the way—with a brief descriptor of how organizations can solve them.

Challenge	Solution
Discouragement from failures	Developing a fault-tolerant culture
Siloed, unprocessed data	Creating a modern data strategy that includes data lakes
Finding the right business problems	Building blended teams that include both technical and domain experts
The machine learning skills gap	Adopting new organizational models, processes, and team management philosophies
Sustainably scaling beyond pilot projects	Leveraging end-to-end tools like Amazon SageMaker to simplify machine learning development
Measuring the results	Forgo traditional ROI metrics in favor of agility, competitive advantage, and risk tolerance; use the value tree model

To learn more about how organizations can overcome obstacles and accelerate their machine learning journeys, visit the AWS machine learning resource hub.

Get started »