Harvard Business Review
ANALYTIC SERVICES

WHITE PAPER

Optimizing Machine Learning-Driven Business Outcomes by Adopting MLOps

Sponsored by
Capital One

SPONSOR PERSPECTIVE

The use of artificial intelligence and machine learning (ML) to drive business transformation and reimagine customer experiences has become ubiquitous across industries and throughout organizations large and small. And one thing has become clear: the ability to deploy ML in and of itself is not a silver bullet for success. Today, an enterprise's ability to leverage ML to its fullest has reached a critical juncture.

While many companies have built strong ML capabilities, far fewer have been able to deploy the majority of their ML models to production, leaving significant value on the table. Scaling ML to realize its maximum potential is a highly methodical process based on a set of standards, tools, and frameworks, broadly known as machine learning operations, or MLOps. MLOps focuses on the entire life cycle of design, implementation, testing, monitoring, and management of ML models and has three primary goals: the first is to develop a highly repeatable process over the end-to-end model life cycle, from feature exploration to model deployment in production. Another goal is to hide the infrastructure complexity from data scientists and analysts so that they can focus on their models and strategies. And the third goal is to develop MLOps in such a way that it scales alongside the number of models and modeling complexity without requiring an army of engineers.

Today, every enterprise serious about embracing machine learning is turning to MLOps. MLOps helps standardize and, to a degree, automate certain processes so engineers and data scientists can spend their time on better optimizing their model parameters and business objectives. MLOps can also provide important frameworks for responsible practices to mitigate bias and risk and enhance governance.

In order keep up with the fast-paced world of ML, organizations would do well to prioritize an MLOps strategy. As experts in this report agree, a targeted strategy can offer a reliable, nimble, and efficient approach to effectively embedding ML in a way that delivers value to the business, its employees, and its customers. But implementing MLOps is not without its challenges. It takes significant time, effort, and resources to develop the infrastructure needed to operationalize ML reliably, and at scale, across the enterprise in a repeatable way.

Capital One has sponsored research by Harvard Business Review Analytic Services to examine the vast and complex landscape of how organizations effectively use ML at scale, with an eye toward understanding the transformative potential of MLOps. Through interviews with ML consultants, analysts, academics, and practitioners, this white paper elucidates the challenges and opportunities that come with MLOps, including how organizations can get the most out of their investments in ML with the right strategies.

Taken together, the insights and best practices offered in this report can set up an enterprise's ML efforts for success by delivering the return on investment—and the differentiated value—that, when done right, ML uniquely makes possible.

Abhijit Bose

Managing Vice President,
Head of Enterprise Machine Learning
and AI Engineering

Capital One

Optimizing Machine Learning-Driven Business Outcomes by Adopting MLOps

Many technology teams release and manage software in a methodical way that improves development efficiency, delivery schedules, and software quality and reduces the chance of outages in production. Machine learning operations (MLOps) intends to do the same for machine learning (ML) models by standardizing and automating application development, deployment, and testing and management. Companies without mature MLOps programs could find their competitors outpacing them in using ML to drive value, increase revenues, hone supply chains, enhance flexibility, and guarantee well-managed processes across the organization.

"ML is going to permeate all these things eventually," says William McKnight, president of McKnight Consulting Group, an information management consulting firm based in Plano, Texas. "If you are not riding the crest of the wave of MLOps, you are going to be lagging behind in ML capabilities." In fact, Needham, Mass.-based technology consulting firm IDC believes that 60% of enterprises will operationalize their ML workflows through MLOps by 2024.

In pursuit of greater business resilience, agility, and performance, organizations are investing considerable time and resources in artificial intelligence (AI) and ML, but heretofore companies have struggled to generate business outcomes such as increased revenues or faster time to market from these technologies. Indeed, 90% of ML models are not deployed to production, according to IDC. When ML is put into production, it can be used in a wide variety of ways, from creating personalized experiences

HIGHLIGHTS

Companies without mature machine learning operations programs will find their **competitors outpacing them in using machine learning** (ML) to drive value, increase revenues, hone supply chains, enhance flexibility, and guarantee well-managed processes across the organization.

ML models make **decisions on many critical matters**, from enabling lenders to confidently approve a mortgage loan to suggesting the best course of medical care. Companies need **transparency** regarding how those decisions are made.

ML is not simply a process of data scientists writing deeply technical code. **Scaling ML is a highly methodical process** based on a set of standards, frameworks, and organizational collaboration.

tailored to each customer to improving fraud detection and enhancing the overall quality of customer service.

At many organizations today, however, data scientists are developing hundreds or even thousands of ML models that never see the light of day. "If you are only able to put a handful of ML models into production, you are leaving a lot of value on the table," says Diego Klabjan, a professor of industrial engineering and management sciences at Northwestern University.

Many things are preventing ML models from being put to work. With an eye to improving their use of ML, companies are exploring MLOps, which drives business outcomes by focusing on the entire life cycle of design, implementation, and management of ML models. Where DevOps creates a set of practices and a culture to improve software development, MLOps takes the next step, applying DevOps tools and approaches to the much more complex areas of ML model development and delivery, allowing those models to be scaled across the organization.

Early adopters of MLOps report up to a tenfold increase in productivity, five times faster model training, and up to a 50% increase in compute utilization, according to research from cnvrg.io, a Santa Clara, Calif.-based AI infrastructure provider.

To effectively implement MLOps, though, companies must address a number of issues, including personnel and technology. This report will explore the unfolding MLOps maturity landscape for organizations and identify the emerging

SPONSOR INSIGHT

The Importance of Platforms

To keep pace with the scale and complexity of today's machine learning (ML) capabilities, many companies leverage sophisticated internal platforms. Enterprise technology platforms enable cross-functional teams to test, launch, and learn at a rapid pace; reduce duplication and standardize capabilities; and provide consistent and integrated experiences, asserts Marcie Apelt, managing vice president, AI, machine learning, and Innovation Lab at Capital One. In short, they help transform technology like ML into a true competitive advantage. A strong cloud platform allows companies to leverage big data to build new ML applications that accelerate, enhance, and deliver on new, more meaningful customer experiences, Apelt adds. When companies combine cloud-native platforms with data at scale, MLOps, or machine learning operations, becomes more manageable—as does the ability to advance and experiment with newer, more innovative ML-driven products and experiences.

best practices for MLOps and the implications this approach has for companies' ML talent, performance, infrastructure, governance, and operations at scale. It will also examine the key barriers to unlocking the value in enterprise-wide ML, as well as the strategies and best practices organizations are leveraging to overcome these challenges.

Wrestling with a World of Data

Every second, 1.7 megabytes of data are created for each person on earth. In a world awash in internet-of-things devices, sensors, connected medical devices, and smart roads, ML is the only way to process and make sense of this tidal wave of information. ML models can discover patterns in huge mounds of data, find anomalies, generate insights, and make predictions and decisions that allow companies to outpace competitors.

"I can't call ML a competitive advantage anymore because everyone is trying to do it," says Neil Sahota, author of the book *Own the A.I. Revolution* and chief innovation officer at the University of California, Irvine. "We're actually generating data for machine consumption rather than human consumption. If you're not prepared to do that, you're already behind the curve."

To date, about a third of companies (36.6%) have at least 50 ML models in production, according to "The Industry Is Ready for Machine Learning Observability At Scale," a survey of 900 ML teams and technical executives produced in December 2021 by Arize, an AI company based in Berkeley, California.

"I can't call ML a competitive advantage anymore because everyone is trying to do it. We're actually generating data for machine consumption rather than human consumption. If you're not prepared to do that, you're already behind the curve," says Neil Sahota, author and chief innovation officer at the University of California, Irvine.

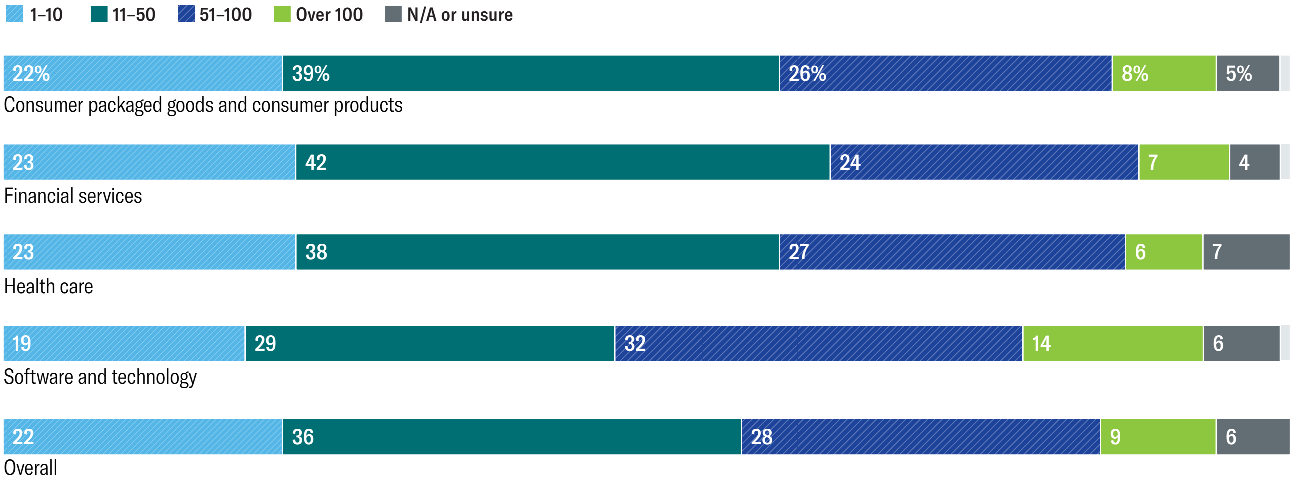
The software and technology (13.7%) and consumer packaged goods industries (8.2%) lead the pack in having the most companies with 100 or more ML models in production. **FIGURE 1**

FIGURE 1

Model Cultures

The technology and consumer products industries have the most machine learning models in production.

Percentage of companies in each industry, by number of machine learning models in production



Source: Harvard Business Review Analytic Services survey, March 2020

Note: Due to rounding, totals may not add up to 100%.

ML models can discover patterns in huge mounds of data, find anomalies, generate insights, and make predictions and decisions that allow companies to outpace competitors.

In explaining the power of ML, Sahota points to a firm focused on mortgage lending and business insurance whose earnings dropped precipitously after those sectors took a major downturn. The company's stock price crashed, its credit rating risk dropped significantly, and it needed to raise capital quickly to cover liabilities. "Normally in that kind of situation, they would start selling off assets to raise capital to cover shortages," he says. "By using ML, they discovered a whole different solution where they could create a new type of product based on the assets they already had, which would start generating large cash flows."

The company developed new products in the large-ticket, higher-risk reinsurance market. At the same time, it used AI to implement operational improvements and trained people rapidly to accelerate organic growth.

That new product generated enough revenue to cover the shortages. Within a year, the company's stock price had risen 117%. Discovering those types of insights from mounds of data that are too complex and voluminous for human brains to absorb is the stock-in-trade of ML and the reason so many organizations are investing heavily in the technology.

Issues with Machine Learning

For all the promise of ML, companies have difficulty leveraging the technology effectively. A mere 8% of 750 business decision makers considered their companies' ML programs sophisticated, according to the "2020 State of Enterprise Machine Learning" survey done in October 2019 by Algorithmia, a division of DataRobot, a Boston-based AI company.

Deployment of ML models also has been slow, the study found. Twenty-two percent of executives said it takes between one and three months to deploy a newly developed ML model into production—where it can deliver business value. Another 18% said it takes more than three months. Such lengthy processes fly in the face of the agility and quick decision making that companies want from ML.

To understand the delays, it's important to appreciate how ML models are developed. ML is not simply a process of data scientists writing deeply technical code. Scaling ML is a highly methodical process based on a set of standards, a framework, and organizational collaboration. Without this rigor, poor data quality can be glaringly apparent when models are put into production. Models can degrade faster without proper maintenance and monitoring, causing them to provide insufficient value or even become detrimental to the organization over time.

During the pandemic, for example, supply chains were disrupted because demand-planning ML models were not updated frequently enough. Companies found themselves with inadequate or excessive inventory because their models

SPONSOR INSIGHT

Establishing Well-Managed, Responsible Machine Learning Practices

As organizations increasingly rely on machine learning (ML) for business-critical decisions, they need to be confident that their ML models and outputs mitigate bias and are fair and responsible, says Sonia Dogra, vice president of technical program management, artificial intelligence, and machine learning at Capital One. To achieve this goal, companies need to proactively build well-managed processes and controls in a very thoughtful, structured way that can accelerate deployment at the same time. Whether an organization is using firsthand research and insights or third-party consultants to accelerate and operationalize responsible ML practices, it takes significant engineering work to bake these processes into an ML infrastructure, she explains. Running applications fully in the cloud can make it much easier to do this, adds Arturo Hinojosa, vice president of product for machine learning platforms at Capital One. If data is scattered throughout multiple systems, an organization cannot uniformly apply controls; it cannot even define common standards and scale. Every project, in every one of those silos, would be very painful. Responsible ML development, like MLOps broadly, should be approached with a holistic and unified method, he recommends.

> “Once you get up and running with a few algorithms, the momentum for using ML is strong because people see the power that the models are delivering to the business.”

William McKnight, president of McKnight Consulting Group

> "MLOps can improve the quality of model production, promote experimentation, and manage regulatory requirements," says Rohit Tandon, global leader, AI services and technology sector leader–data, analytics, and AI for Deloitte Consulting.

were operating on data and assumptions that were suddenly no longer applicable in the changing business environment.

ML models make decisions on many critical matters, from enabling lenders to confidently approve a mortgage loan to suggesting the best course of medical care. As a result, companies need transparency regarding how those decisions were made.

Sahota recalls a colleague who developed an algorithm for a simple image recognition task—distinguishing a picture of a dog from a picture of a wolf. However, the algorithm started labeling as dogs pictures that were clearly wolves. After a while, the data scientists realized in their training set of photos that were used to develop the model, every photo of a wolf had snow in the background. So essentially, the "wolf detector" model had been trained to recognize snow, and if there wasn't any, the wolves were identified as dogs instead.

"Fast-forward a couple of years and we've had problems with recognition tools that were really good at recognizing white males but not other ethnicities and genders," Sahota says. This circumstance wasn't a case of intentional bias or discrimination. Rather, Sahota explains, it was a situation of "just training the models by using pictures of celebrities and people that looked like them, which introduced implicit bias into the system."

MLOps can head off such problems by providing a framework that explicitly checks for such matters. A company that effectively deploys MLOps excels in these five areas, starting with deploying ML models consistently. It uses standardized, consistent processes and controls to monitor models in production for drift, as well as for data and feature quality. It replicates and retrains ML models with confidence in production. Also, it nurtures the ability to take ML models developed by data scientists and bring them all the way through the company's quality assurance processes and into production without the burden of much manual or engineering work in between those steps. Lastly, it ensures that its ML infrastructure is resilient, consistently scanned for vulnerabilities, and well managed across the board.

Implementing MLOps can have a huge impact on a company's effective use of AI. Indeed, a Deloitte Consulting survey of 2,875 technology and business executives in 11 countries in May 2021 found that more than a quarter (28%) of the most advanced users of AI—a group that the consultancy calls "transformers"—have implemented MLOps. Less than a fifth (17%) of "underachievers," companies that haven't adopted enough leading practices to help them effectively achieve more meaningful outcomes, can make the same claim. **FIGURE 2**

Stages of MLOps Maturity

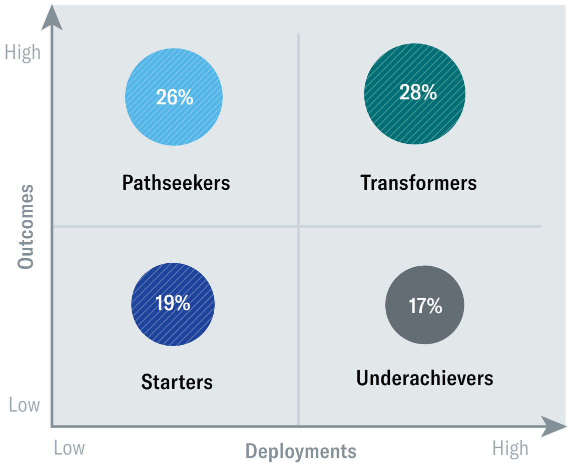
MLOps applies DevOps tools and approaches in order to industrialize and scale machine learning, from development and deployment to ongoing model maintenance and management.

As a result, "MLOps can improve the quality of model production, promote experimentation, and manage regulatory requirements," says Rohit Tandon, global leader, AI services and technology sector leader–data, analytics, and AI for Deloitte Consulting.

FIGURE 2

Transforming through Machine Learning Operations

The most advanced users of artificial intelligence are implementing machine learning operations practices.



Source: Deloitte Consulting, October 2021

Developing ML models is a complex, highly iterative process that includes preparing data, training the model, validating the results of the model and releasing it, and monitoring the model for any performance or accuracy flaws after release.

"Once you get up and running with a few algorithms, the momentum for using ML is strong because people see the power that the models are delivering to the business," McKnight Consulting's McKnight says. "But you need a way to organize the process so the models aren't developed within one department and never shared."

McKnight has developed an MLOps maturity model denoted by five stages that determine where an organization stands in its use of the technology.

The model is based on five criteria that form the foundation of MLOps and a company's use of ML. This foundation includes how well MLOps aligns with an organization's culture and priorities; whether the technology architecture can manage data and deploy models coherently; whether the company has sufficient data science skills and domain expertise; whether processes allow the effective, efficient, and measurable delivery of activities necessary for ML models; and whether proper governance is in place to enable trust in and to explain the ML models.

Given that ML itself is in an early stage among organizations, it's not surprising that MLOps' development is embryonic. McKnight says, in terms of maturity, most companies are in the middle. They have not yet engaged with ML in a strategic way or have made only some progress in delivering the benefits of ML. **FIGURE 3**

"There is clearly a long way to go," he says. "The MLOps model is not set in stone by any means. However, in a few years, if you're not a four or five out of five in maturity, you might find that you are feeling the competitive pressure to pursue a machine learning strategy."

Promoting Trust in ML

MLOps can establish and enforce program-level guardrails to promote accountability. The MLOps framework enables deployment and development teams to stay well managed and responsive, mitigate bias, and adhere to privacy and security regulations, governance, and compliance more easily.

"MLOps streamlines the process of updating these models as the business evolves," says Luis Ceze, a professor of computer science and engineering at the University of Washington. "As you learn more, you create best practices for curating and keeping your experiments and training data, validating the models, and determining when they are ready to transfer into the production phase."

MLOps can address the common problem of "model drift," where changes in the real-world environment cause the model's predictive abilities to decay over time.

FIGURE 3

Five Stages of Machine Learning Operations Maturity

Machine learning operations (MLOps) is a journey from manual processes to an explainable, transparent strategy. Here are the five stages that companies go through on their journey of implementing and using MLOps.

Stage	What It Means
1	Just gaining an understanding of how to use machine learning (ML). No data scientists hired. Early data models built without much success. There is a belief that whatever DevOps processes are in place will handle machine learning.
2	The data architecture serves most data that would be necessary for ML. A cloud commitment and direction are present, providing scale for ML. First data scientist is hired, and prototyping is done. Full life cycle ML is accomplished with manual processes. MLOps is still an afterthought.
3	This company is actively looking to deliver the benefits of ML across the company. There is recognition of ML at the executive level. However, early processes in use resemble DevOps and will not scale. The company begins making a complete copy of its ML repository, the collection of databases, domain theories, and data generators.
4	There is company-wide embracement of ML. Benefits have been produced and realized. There are numerous and ample data scientists, and the data architecture has matured so that more ML benefits can be realized. Although processes still aren't fully consistent, the company has embraced MLOps and is rapidly adapting it.
5	The business has fundamentally changed due to ML, and it could not have done so without MLOps. ML is applied to initiatives wherever possible. MLOps is nurtured as much as ML and includes model sharing, reusability and reproducibility, model diagnostics, and a strong path to production. Governance has become central to ML strategy, ensuring outcomes that are explainable and transparent.

Source: Harvard Business Review Analytic Services survey, April 2020

> "The question becomes providing continuous integration and delivery in a transparent fashion, synchronizing changes in software in development with software that is already being used by customers," says Northwestern University's Klabjan.

"The fun really starts once you put the model in production, because then you have to monitor it, tweak it, and detect changes or drift," Northwestern's Klabjan says. "Ideally, you automated the whole process. The question becomes providing continuous integration and delivery in a transparent fashion, synchronizing changes in software in development with software that is already being used by customers."

MLOps tools can automate the collection of information to determine how models were used and whether the data was handled properly. Such steps can promote trust in ML, making people throughout the organization more ready to embrace the technology and making it easier for workers to understand ML and explain it to others.

"ML has been mystified, like it's different than any other type of software development," University of Washington's Ceze says. "The model creation process is different than writing software, but the industry should demystify ML for people so they are as comfortable using it as any other piece of software."

Makeup of MLOps Teams

Most ML teams have between 11 and 50 employees, according to the Arize study. However, 18% of ML teams in the software and technology sectors have more than 100 people, the study found. Financial services and consumer packaged goods companies also tend to have larger ML teams.

The AI/ML landscape requires a wide range of technical personas—data scientists, data engineers, developers, ML engineers, technology personnel, and design and product management experts—who are responsible for delivering and managing the infrastructure to support all the different phases of the ML model life cycle. At this early stage, there is no standard approach to what an MLOps team looks like.

"The companies that have been most successful with MLOps have taken a step back and put together teams which are a microcosm of their entire business," Deloitte's Tandon says. "They have domain knowledge, ML skills, cloud engineering skills, and data engineering skills. Some clients who work in an extremely centralized fashion might have centralized MLOps teams. Some use a spoke-and-hub model. The most important thing is to realize when you start this journey, it's a journey of scale. If you try to do it in bits and pieces, you'll never have enough momentum around it."

Sometimes companies emphasize having someone with esoteric skills on an ML team, like a computer vision expert, who is not core to their operations and is not needed in-house on an ongoing basis. At the same time, organizations can underestimate the need for data engineering and cloud engineering skills in ML team members. "Domain expertise

SPONSOR INSIGHT

Machine Learning and the Quality of Data

It's no secret that machine learning (ML) is driven by data. Much like the broader world around us, data changes constantly, which has big implications for operations run by ML, explains Nurtekin Savas, vice president of machine learning and data science at Capital One. The quality of an organization's data directly impacts the effectiveness, accuracy, and overall impact of ML models' outputs. High-quality data makes organizations' ML efforts more resilient, adaptive, and secure, he notes. It offers the agility to react to data drift in real time and send it back into the model so it can relearn and adjust its outputs accordingly. To be adaptive and resilient in this way, and operate at scale, requires organizations to adopt sound frameworks, tools, data patterns, and governance practices. This includes ensuring data scientists and ML engineers have standard tools, processes, and platforms; making sure data is secure, structured, standardized, and accessible; automating model monitoring and training processes; establishing well-managed, human-centered processes like model governance, risk controls, peer review, and bias mitigation; and creating foundational architectures and frameworks, Savas asserts.

“ML has been mystified, like it’s different than any other type of software development. The model creation process is different than writing software, but the industry should demystify ML for people so they are as comfortable using it as any other piece of software.”

Luis Ceze, a professor of computer science and engineering at the University of Washington

is critical," Tandon says. "It doesn't matter what technology you use if you don't know the business challenge you are trying to solve."

He says companies should bring AI practitioners and data scientists together in one practice while also investing in preconfigured solutions. "Business and domain experts can build use cases around signature issues, data science experts can drive innovation in ML models, and data and ML engineers can use auto-ML tools to stitch together quick ML models," he says.

Technology Underpinnings

Companies can be challenged by a lack of hardware designed for ML use cases. Enterprise Strategy Group, a technology analyst firm in Milford, Mass., reports that "86% of organizations identified at least one of the following areas as a weak link in their AI infrastructure stack: GPU [graphics processing unit] processing, CPU [central processing unit], data storage, networking, resource sharing, or integrated development environments."

"The cloud offers services and tools to get you started with machine learning," says Mike Leone, senior analyst, Enterprise Strategy Group. "If you don't have data scientists, the cloud can provide pretrained models, an environment to upload your data, and off you go. You'll have an app in production right away, which is super appealing for a lot of organizations that don't have all the skills for ML development in-house."

There is a nascent availability of MLOps tools like training libraries, which are a compilation of functions and routines that save developers from having to write the same code over and over. "There are platforms to codify the data and help manage the algorithm," Sahota says, "But they are very basic. They're not like the robust tool set we've developed for DevOps."

IDC reports that 28% of AI/ML projects fail due to lack of necessary expertise, production-ready data, and an integrated development environment, which is an application that allows developers to consolidate the different aspects of writing a computer program.

"You want to enable those developers that are actually doing [this] AI model building, but you need that to be in line with the infrastructure you're using," Enterprise Strategy Group's Leone says. "So having an integrated stack is very valuable."

Klabjan notes that data scientists can be pressed by their companies to evaluate MLOps technology in a compressed time frame. "They'll be told, 'Here's a vendor we're considering, here's a solution we are considering, and you have one week to evaluate it,'" he says. "However, this technology can be fundamental to the future of their ML models, which can be fundamental to the future of their companies, so they should give their data scientists more time to evaluate the technology."

"Is it the ability to take out interference from someone who is interacting with you on your website so you can properly recommend what they should buy? Is it the conversion rate of recommended products? The proper metrics are going to be different for every industry, but establishing the right KPIs is critical," says William McKnight of McKnight Consulting Group.

Best Practices in MLOps

As companies begin to develop MLOps functions, best practices are emerging to encourage the most successful use of ML.

Determine sound KPIs. ML projects should be aligned with business priorities. "Companies set themselves up for failure when a CXO [chief experience officer] declares ML is a mandate and they say, 'Well, OK, I guess I'll try it with this use case,'" McKnight says. "You need to determine the right metrics for success for ML. Is it the ability to take out interference from someone who is interacting with you on your website so you can properly recommend what they should buy? Is it the conversion rate of recommended products? The proper metrics are going to be different for every industry, but establishing the right KPIs is critical."

Publicize your success. As much as ML can contribute to business value, it can also be plumbing that is not noticed by enough business decision makers. For that reason, in internal communications, Sahota believes in making clear the impact of ML's ability to drive value, including by having marketing slogans such as "This business success brought to you by Algorithm X."

Educate executives. To gain support for MLOps, Klabjan says, business executives often must be convinced to go beyond their reliance on and comfort with spreadsheets to track information. "Decision makers need to be aware of how complex the MLOps process is, and keeping track of models and data can't be done on spreadsheets but requires more sophisticated tools," he says.

Focus on innovation. MLOps pulls together a range of skills and relies on automation, workflows, and systems to drive impact on a sustained basis. Process-centricity with MLOps can sometimes obscure the fact that innovation is at the core of using AI and ML successfully. "The MLOps framework should encourage innovation and ensure the ML stays relevant and can be used in the future," Tandon says.

Embrace transparency. McKnight says MLOps should incorporate governance and security frameworks from the outset. "Even if they have not perfected a model, ML operators should take measures to eliminate model bias and the potential for malicious use of ML," he says.

SPONSOR INSIGHT

Collaboration Is Key to MLOps Success

Deliberate, thoughtful collaboration and communication is paramount to any cross-functional technology project, asserts Miriam Friedel, senior director of machine learning engineering at Capital One. When it comes to implementing MLOps, or machine learning operations, especially in large organizations, clear standards and automated processes can help foster collaboration, break down silos, and improve overall performance. She advises companies to create intentional technical contracts to help standardize points of connection to ease coordination across distributed systems and ensure that resiliency isn't solely dependent on human communication. Businesses should also create an ML infrastructure that provides visibility into how models are performing at scale and recognizes when models begin to change. Achieving this visibility and collaboration requires that companies follow best practices. Friedel says executives should prioritize a clear set of use cases and identify stakeholders and then build for that community first. Next, they should establish shared standards and a unified tech stack and incentivize the reuse of tools where possible while focusing on monitoring and logging. Finally, organizations should create solutions that the whole team can take pride in building together.

> "There are retailers out there with hundreds of models that are deployed in a modular way, where they are all interconnected and working together. How can you expect someone to manage all that without a framework like MLOps?" says Mike Leone, senior analyst, Enterprise Strategy Group.

Conclusion

As companies look to put dozens—or even hundreds—of ML models into production, they can benefit from the same discipline that software development gained from DevOps.

"There are retailers out there with hundreds of models that are deployed in a modular way, where they are all interconnected and working together," Leone says. "How can you expect someone to manage all that without a framework like MLOps?"

Still, MLOps is a function not just of the number of models that a company has in production but also of the importance of those models, Ceze says. "I can think of multiple companies that have a single model that is so central to their business and delivers so much business value, and they need MLOps to support it." For example, a virtual assistant in the customer service department that is serving millions of customers could be so essential that a company should have MLOps in place just for that feature.

Organizations that have been successful at scaling AI and ML have implemented a set of standards and developed a framework to build production-capable AI and ML building blocks. MLOps will be essential to enabling and demonstrating the successful use of ML, which will be integral to companies' future success.

"Ten years from now, there's not going to be a product or service that doesn't have some sort of ML component to it," Sahota says. "If you're not doing ML well, you're never going to reach the optimization you need to compete. You've got to get good at ML—period—which means you have to get good at MLOps. And the earlier you do that, the more you're setting yourself up for success."

SPONSOR INSIGHT

Best Practices for MLOps Success

At a high level, setting executives up for success with machine learning operations, or MLOps, comes down to a few best practices, says Zach Hanif, vice president of machine learning at Capital One. First, companies should ensure that their machine learning (ML) compute environment can handle complex compute solutions at scale (moving to the cloud can make this more cost-effective, flexible, and efficient). Next, businesses should advance ML functionality so they can automate model monitoring and training and confirm that it's performing as intended as they push into production. Hanif advises companies to identify where they can reduce the lag between production and analytical data environments. Ali Rodell, Capital One's senior director of machine learning engineering, adds that organizations should examine ways to create more visibility into core cloud infrastructure across the enterprise, including continuous integration and delivery management, container deployment, security management, and governance. In addition to paving the way toward a smoother MLOps infrastructure, he says, these considerations can allow DevOps engineers and site reliability engineers to guarantee consistent reproducibility, model monitoring, and maintenance of ML systems.

Harvard Business Review
ANALYTIC SERVICES

ABOUT US

Harvard Business Review Analytic Services is an independent commercial research unit within Harvard Business Review Group, conducting research and comparative analysis on important management challenges and emerging business opportunities. Seeking to provide business intelligence and peer-group insight, each report is published based on the findings of original quantitative and/or qualitative research and analysis. Quantitative surveys are conducted with the HBR Advisory Council, HBR's global research panel, and qualitative research is conducted with senior business executives and subject matter experts from within and beyond the *Harvard Business Review* author community. Email us at hbranalyticservices@hbr.org.

hbr.org/hbr-analytic-services

Copyright © 2022 Harvard Business School Publishing.

MCCRE3284