EXECUTIVE BRIEFING

The Business Value of MLOps

Thomas H. Davenport

President's Distinguished Professor of Information Technology and Management
at Babson College, a Visiting Professor at Oxford University's Saïd Business School,
a Fellow of the MIT Center for Digital Business, and a Senior Advisor to Deloitte's AI practice.

DataRobot

Thomas H. Davenport

Thomas H. Davenport teaches analytics and big data in executive programs at Babson, Harvard Business School, Harvard School of Public Health, and the MIT Sloan School. He pioneered the concept of “competing on analytics” with his best-selling 2006 Harvard Business Review article (and his 2007 book by the same name). His most recent book is The AI Advantage from MIT Press.

He has written or edited over twenty books and over 300 articles for Harvard Business Review, MIT Sloan Management Review, the Financial Times, and many other publications. He writes columns for Forbes, MIT Sloan Management Review, and the Wall Street Journal. He’s also been a LinkedIn Top Voice for both the education and tech sectors.

Machine learning has been used by companies for several decades now, but over the last few years it has become a critical business resource for many organisations. The capabilities to train models with data, to generate accurate predictions about important business outcomes, and to analyze both structured and unstructured data have made machine learning the most important component of artificial intelligence.

Machine learning has been made even more powerful and useful by the advent of automated machine learning (AutoML), which both enables quantitatively-oriented amateurs to generate high-quality models and improves the productivity of professional data scientists. Model creation was once the limiting factor in machine learning success but with AutoML that is much less a constraint.

However, as machine learning models are increasingly put into production deployment and used to make critical business decisions, the primary challenge becomes operation and management of multiple models. Machine learning operations (MLOps), is the technical response to that issue. These tools help place machine learning models into production deployment and then log and monitor the performance of models in production to ensure that the models are performing well. MLOps tools also document key aspects of machine learning models so that those who did not develop them can understand and revise them after development.

In that sense, the rise of MLOps is a response to a problem arising from success. It wouldn’t be necessary if machine learning models weren’t being created in greater numbers, and if they weren’t being used extensively for predictions and decision-making. MLOps is based on the realization that machine learning models are a valuable business asset that should be managed as such.

The COVID-19 pandemic illustrated the value of MLOps for Lifeblood's operations. Many of the unit's models address the important issue of blood collection forecasting, which was heavily impacted by pandemic lockdowns. Lifeblood needed to monitor existing models daily for drift and retrained many models every few weeks. MLOps also helped them deploy different types of models at a fast cadence that were more responsive to the short-term blood donation trends.

MLOps is an extension of the DevOps idea — an effort to bring operational discipline to system development. Just as DevOps has the objective of improving system development productivity and quality, MLOps is intended to improve the speed and success rate with which machine learning models are deployed into production and operated over time. Since model deployment into production is how machine learning creates economic value, MLOps is a means of achieving greater economic return from this form of AI.

This report focuses on the benefits of MLOps tools and processes for different types of organizations. It's based on interviews with MLOps user companies and several MLOps experts. Some were from ParallelM, which was founded in 2014 and was the first MLOps software provider. DataRobot acquired the company in 2019. The MLOps customer companies that I interviewed were users of DataRobot's MLOps solution. Many of the interviews took place during the COVID-19 pandemic, and most customers reflected on the impact of the pandemic on their models and MLOps.

Why MLOps?

The organizations I interviewed had various reasons for adopting an MLOps solution. The Australian Red Cross Lifeblood, for example, adopted an end-to-end perspective from the beginning of their data science focus three years ago. Andrew Clarke, who heads the data science group there, said that "The MLOps approach to machine learning and AI is how we operate. We knew we needed the ability to rapidly drive deployment, to compare models, and to track our models that we might want to put into production. We wanted to avoid technical debt in the machine learning area." Lifeblood, like many users of MLOps, also uses AutoML software (provided by DataRobot). The organization has many analysts who can create models using AutoML, but relatively few data scientists who can monitor models for drift and lessened performance.

On the other hand, Norfolk Iron & Metal, a medium-sized steel processing and distribution company in the Midwestern US, approached MLOps not from a data science perspective, but from one of improving business decision-making. Neither Ben Dubois nor Mike Bamsey, the two leaders of the MLOps work, could be described as data scientists. Bamsey is the company's executive vice president, and Dubois is an operations analyst.

In part at Bamsey's instigation, the company has made a shift in recent years from experience-based to data-driven decisions, particularly for repetitive decisions like estimating how long jobs will take and pricing them, and how the settings that go into a shaping and cutting machine could reduce scrap. They read about and attended seminars on machine learning, and concluded that it could improve many of their decisions. They initially adopted AutoML and tried to fold it into as many workflows as possible.

After Bamsey and Dubois began to proliferate models, they concluded that the company should manage them with MLOps. Norfolk Iron & Metal had only two major use cases in production, but there were multiple models within each of them. The company has more models ready to go, but it has not yet put them into production.

They were interested in MLOps because they simply wanted to ensure that their models in production stayed current and did what they wanted them to do. They now monitor the performance of the models and results from them, and retrain them when necessary. Their business has also changed significantly because of COVID-19; they saw a step change in demand for their metal products, and they have integrated AutoML models into their demand and inventory planning system. MLOps alerted them to the need for retraining their models to understand and predict the surge in demand.

Another completely different reason for adopting MLOps is represented by the banking industry. In that industry, the management of machine learning models — particularly those for credit or risk decisions — are heavily regulated. Machine learning models are nothing new to the industry, in that neural network models were used for credit authorization and fraud detection in the 1980s and have been used continually since then. Regulators, such as the Federal Reserve, have issued regulations (specifically SR 11-7) mandating certain approaches to "model risk management," which at the time of issue (2011) was only possible with manual processes. Almost all large banks doing business in the US have established such processes with large numbers of human workers to perform and support them.

Now, however, many of the model risk management processes can be performed by MLOps software. Some banks are using MLOps tools to automate aspects of model risk management. At others, there is still a reluctance to replace the manually-performed tasks in part for concern about regulatory acceptance. And all banks are reluctant to speak on the record about MLOps given the regulatory sensitivity of the issue.

However, I spoke with several experts in banking MLOps, and they confirmed that many leading banks are exploring or already using MLOps for credit and risk model management. These experts are particularly interested in deployment approval workflows that are built into MLOps systems, in which stakeholders (using role-based access) can sign off on models throughout their lifecycle. Also of appeal is the idea of risk-weighted controls, which are based on the notion that some models are more critical than others to a bank's risk exposure. With this capability, banks can manage models differently based on their materiality.

It's clear that large banks will eventually need to adopt MLOps, but thus far only the most technologically progressive are doing so. At some point regulators may demand systems-based monitoring of production models at the enterprise level, which would hasten adoption.

Who Is Using MLOps?

The industries most focused on MLOps are generally those with the most data and the most need to make predictions with machine learning models as a component of their businesses. Financial services are at the top of that list, and the industry will eventually move from manual model management to automated approaches. One company that is already doing so is Net Pay Advance (NPA), a payday and installment loan lender in several US states. Clayton Howard, head of data science for NPA, said that machine learning is used throughout the company's operations — marketing, in-house collections, and customer relationship management — but the single most important use case is credit underwriting. He estimates that the company has about 50 models in production.

When the company realized the success it was having with decisions based on machine learning, executives realized that the models — particularly those used in credit underwriting — were a valuable asset that helped them perform better financially. There are as yet no regulatory requirements to manage the models across the enterprise, but data sources are regulated, and Howard believes it makes regulators more comfortable to know what models are using what data. Before MLOps, Howard himself was the primary person monitoring the models; having the DataRobot MLOps solution is like having 20 of him, he commented.

The primary value to NPA from monitoring models is determining whether the models exhibit drift, which lets the company know they need to refresh or retrain them. NPA began to use MLOps just as the COVID-19 pandemic took off when there were rapid changes in demand for credit and data about consumer creditworthiness such as employment status and income. The NPA data scientists weren't initially sure how to react, but their MLOps system tipped them off that several credit attributes were showing more drift. They immediately began to shorten the timeframe of the data they used to train their models.

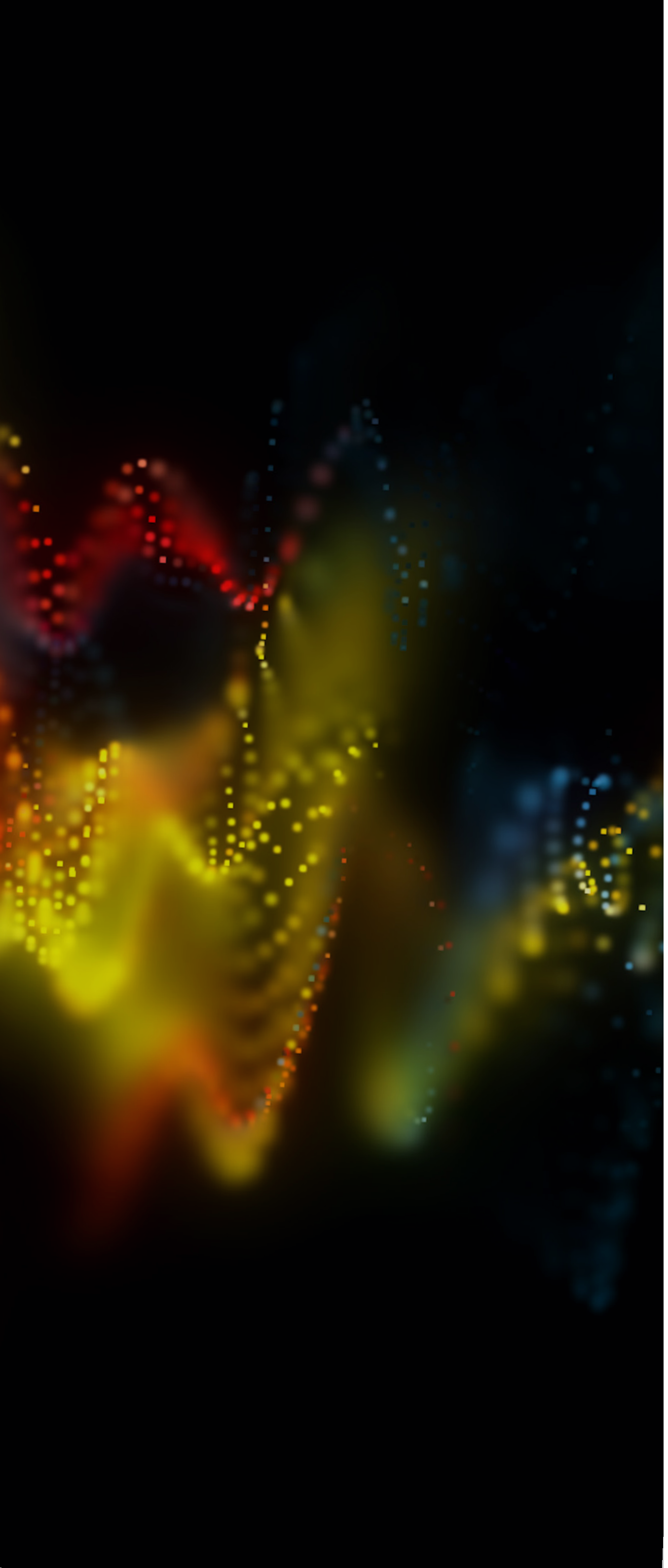
Howard commented that the most valuable aspect of MLOps is looking for drift, but simply being able to keep track of models is also valuable. He and his team looked at how often the models are used, how many predictions the models made, whether and why they fail, and so forth. He summarized

"MLOps is a security blanket for our important machine learning assets."

Insurance is another data-rich industry that has been using analytics in key processes for decades. The industry has historically made extensive use of rule engines in applications for underwriting and fraud prevention, but it's increasingly moving to machine learning for analysis of more data and more precise predictions. One company in that industry uses rules, machine learning, and MLOps to increase claims payment accuracy. Decisions about which claims need to be audited were previously made only with millions of rules, but the company now supplements the rules with machine learning models. Several million automated predictions about claims needing to be audited are made each day.

The Operational AI group at the company applies AI techniques to enhance the quality and productivity of operational processes. It is comprised of a centralized team of data scientists, AI solution architects, engineers and data analysts. They drive the development of a suite of production AI solutions on a robust, scalable, sustainable AI platform with measurable return on investment. All of the group's machine learning models, which generate scores predicting whether a claim is worth auditing, are generated with AutoML. For the last six months, the team has also been using MLOps.

The group has between 25 and 30 models in production. MLOps was adopted to ensure that models are robust and would run, and provide scalability and sustainability over time. Their competitive advantage, the group's head argues, derives from building high-quality models faster than their competitors and keeping them up-to-date over time. Before MLOps, a small team spent a lot of time monitoring and maintaining a smaller number of models; now it takes less effort to manage more than double the number of models. The models are developed by a development team, which builds them and gets them up and running. Then they turn the models over to the MLOps team, which found it easy to get the models into the MLOps system and ensure their ongoing effectiveness at finding claims to audit.



Other industries that are finding MLOps systems useful are:

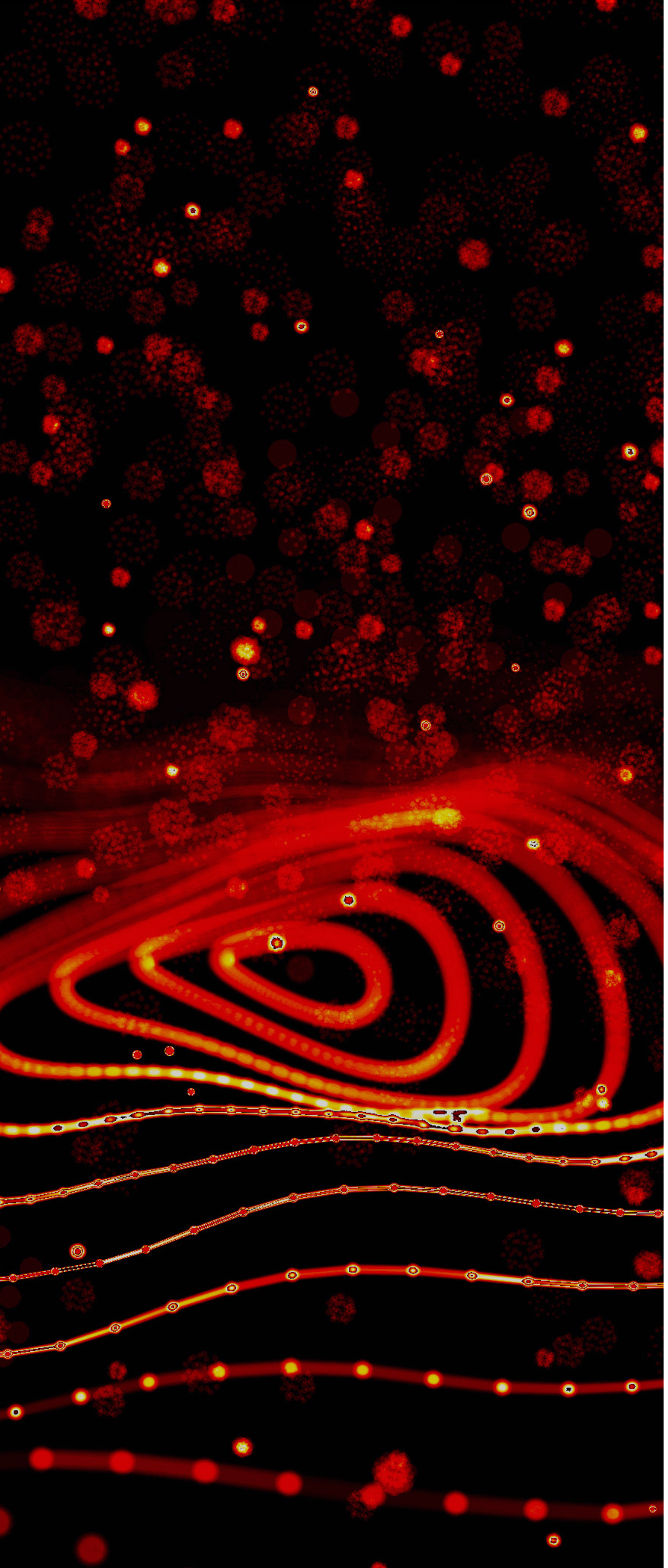
Telecommunications companies, which are managing network operations and customer churn models.

Manufacturing companies, which are using machine learning models to predict asset maintenance needs and identify quality problems.

Marketing and advertising organizations, which have multiple machine learning models in production to present targeted digital ads to consumers.

Companies with platform-based business models, which use machine learning extensively to match customer needs and supplier offerings.

The healthcare industry will undoubtedly make extensive use of diagnosis and treatment-related machine learning models at some point, and such models are increasingly emerging from research programs. As yet, however, these machine learning models are not generally used in clinical practice. When they are, it's likely that both good management practice and external regulation will force the use of model management tools like MLOps.



Determining the Benefits Achieved from MLOps

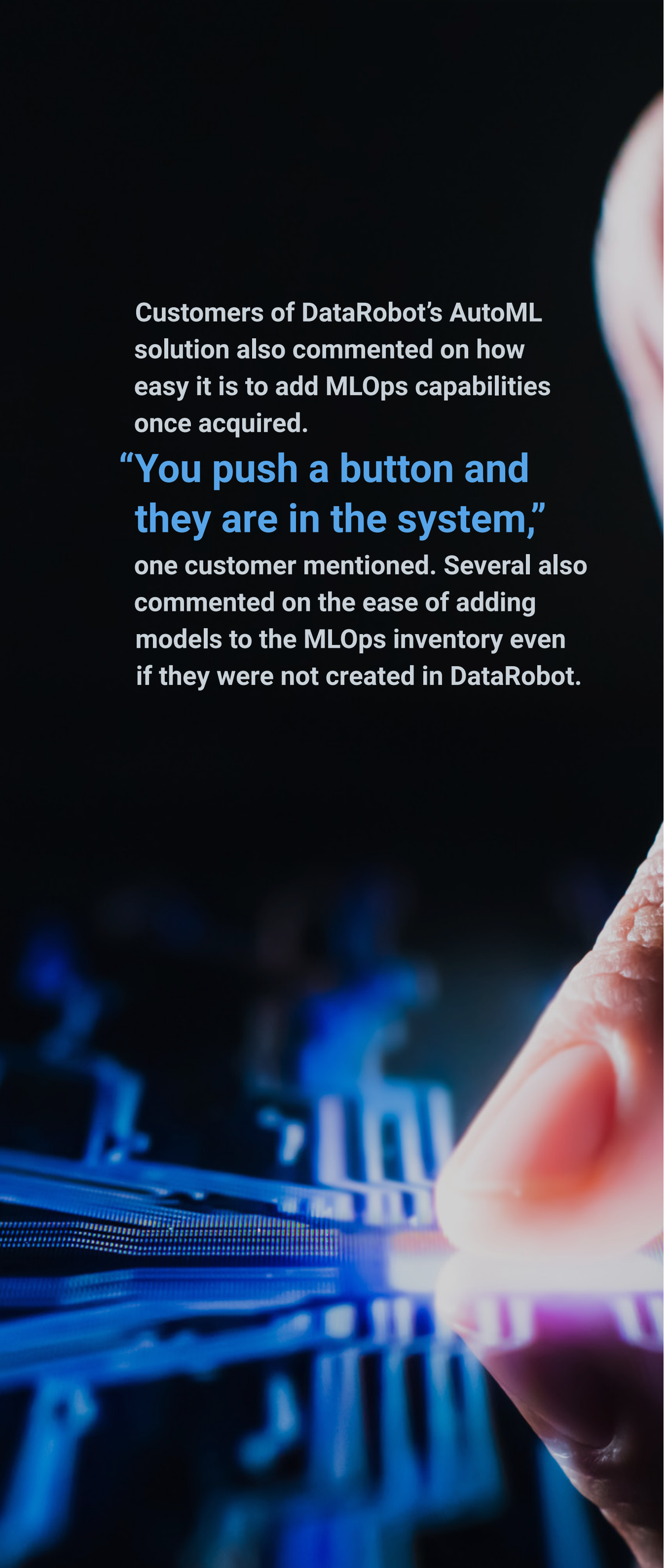
Measuring the benefits from MLOps can be difficult if an organization isn't already doing the same tasks manually. MLOps is fundamentally insurance (or as Clayton Howard put it, a "security blanket") against models drifting and going awry or against the creators of the models leaving and no one having any idea how the model works. Since a company can't really buy that type of insurance from an insurance company, it's hard to compare the price or value of MLOps to its benefits.

If a company's operations are highly dependent upon machine learning models and it has established manual processes for monitoring and checking them, the company is likely to find that MLOps will enable a major productivity improvement as well as higher-quality outcomes. That benefit was mentioned by several companies I interviewed in that situation. Higher productivity means that a company needs to hire fewer expensive data scientists to manage models through their lifecycle.

More rapid productionizing of models may also yield substantial financial benefits. If, for example, a bank has credit models that yield greater profitability, it should put them in place as soon as possible in order to maximize returns. Without an MLOps process in place, many models don't make it into production quickly or at all and thus yield lesser returns or become sunk costs.

If a company is highly dependent upon machine learning models and has no manual management approaches in place, its downside risk is very high. In most cases it should be possible to estimate the cost to the business of using a bad model or two for an extended period of time. In these cases, the costs of an MLOps solution will be well below those estimates.

If a company is highly regulated with regard to its machine learning models (as are banks), it's likely that MLOps will eventually be viewed as a mandatory aspect of doing business. A company may be able to implement MLOps ahead of regulatory mandates and save money compared to the manual processes currently required by regulators.



Customers of DataRobot's AutoML solution also commented on how easy it is to add MLOps capabilities once acquired.

"You push a button and they are in the system," one customer mentioned. Several also commented on the ease of adding models to the MLOps inventory even if they were not created in DataRobot.

MLOps and AutoML

Automated machine learning (AutoML) makes MLOps both easier and more necessary to implement. Companies using AutoML can create many more models in a given period of time and by a given number of data scientists. But another benefit from AutoML adoption is the democratization of model creation that it makes possible. Large numbers of quantitatively-oriented analysts can produce models for exploration or production. This democratization, however, makes it even more important to ensure that models are carefully managed in the deployment process and thereafter. In short, it's not an accident that an AutoML provider like DataRobot would find adding MLOps capabilities to its offerings a good idea.

Organizational Issues Affecting MLOps Implementation

Although the MLOps users I spoke with were obviously able to overcome them, users and experts mentioned several organizational challenges relative to MLOps of which companies should be aware. One is the issue of existing manual processes to manage models that I mentioned are being employed at some organizations in the banking industry. While the presence of manual processes and groups to perform them does indicate that a company is interested in model management, it may also lead to some resistance by human employees who feel that MLOps might threaten their jobs. It's too soon to know whether such fears are justifiable or not, but executives could prevent such opposition by declaring that a model risk management analyst or related type of worker would not lose their jobs to an MLOps system.

Another organizational challenge at some companies is a gap that sometimes exists between IT operations personnel and the data science groups that create new models. Data scientists may feel that deployment and monitoring of models isn't their job; they may argue that they are only responsible for model creation. And IT operations people may be reluctant to deploy, certify, maintain, or otherwise manage data science models. They could feel, for example, that they don't understand the data science models and it's not their jobs to know whether a model performs as advertised. This situation could lead to data scientists throwing their models over a wall to IT operations, and no one on that side of the wall being interested in picking them up. It may also lead to no clear owner for an MLOps solution, even though it could help to bridge the gap.

It's becoming clear that new roles and groups will be necessary to oversee machine learning model operations and ongoing management. Such groups are somewhat rare today, although I mentioned that a small one was created at the insurance company. However, I suspect that many companies will have "Machine Learning Operations" groups sometime soon. Job boards are replete with "DevOps Engineer" roles now, and similar ones will need to be created for machine learning.

The Likely Future of MLOps

It seems likely that as companies grow more dependent on data, analytics, and machine learning, that they will absolutely need monitoring and management tools for these critical business assets. It's also obvious that unanticipated external conditions, such as what we recently saw with the pandemic, the 2008 housing crisis, and climate change, will drive the need and adoption of MLOps as well.

AutoML and MLOps programs, now already somewhat integrated, will perhaps combine into a fully blended Machine Learning Management system.

It also seems likely that such an integrated system would perform more functions with respect to data and models than any AutoML or MLOps offering currently does. They might include, for example, the following system functions:

- Continuous learning systems that can, when significant drift is detected, fetch new training data, engineer the features and train the models, put them into production, and register them in the MLOps library.
- Data or algorithmic bias detection systems that can automatically detect biased data or outcome predictions and seek alternative data to eliminate it.
- AutoML and MLOps systems that can enable the automated generation of world-class challenger models to continuously improve an organization's decision-making; the process tests a variety of alternative models to constantly stress and possibly replace the current champion model in production.
- Assumption monitoring systems that can constantly monitor key data sources in the external environment that might indicate that machine learning predictions from a model are no longer valid.
- Model explanation systems that can, using various explanatory techniques, bring transparency for human reviewers to even the most complex models.
- Greater integration between machine learning management systems and enterprise data catalogs allow model development to draw on new data sources automatically.

With more modeling and monitoring functions performed by automated systems, human data scientists would be left with checking models for reasonableness, ensuring that models are deployed successfully, viewing drift diagnosis and recommendations, and working with stakeholders to frame decisions correctly and take appropriate actions on the basis of model outcomes. Humans will continue for the foreseeable future to have better insight than machines into "big picture" issues that may affect models and their usefulness in addressing business problems. Even with much more automated model creation, deployment, and monitoring processes, there should be plenty of useful things for data scientists and data-driven decision-makers to do.

About DataRobot

DataRobot AI Cloud is the next generation of AI. The unified platform is built for all data types, all users, and all environments to deliver critical business insights for every organization. DataRobot is trusted by global customers across industries and verticals, including a third of the Fortune 50. For more information, visit http://www.datarobot.com/.