Publication date:
28 Sep 2020
Author:
Bradley Shimmin
Distinguished Analyst

Fundamentals of MLOps

Brought to you by Informa Tech

Table of Contents:

Table of Figures:

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Summary

Catalyst

The application of artificial intelligence (AI) in the enterprise has now matured to a point where companies have at their disposal a rich assortment of platforms and technologies designed to simplify and speed the creation of predictive machine learning (ML) business outcomes. Readily accessible and increasingly automated ML platforms have greatly democratized the ML lifecycle, enabling a wider range of companies to invest in ML outcomes without having to invest heavily in data science expertise. And yet bespoke AI development in the enterprise has not yet reached a significant level of adoption. Many companies are experimenting with building ML solutions, but few have put these experiments into production across the enterprise at scale.

The problem stems from an initial focus among enterprise practitioners focused on closing a data science skills gap while ignoring other key aspects of the ML lifecycle, particularly operational issues such as ML model testing, deployment, and monitoring. In response, the vendor market has pivoted to directly target operational issues through ML operations (MLOps) platforms, incorporating numerous DevOps technologies and practices as a way of helping companies scale ML beyond the limits of experimentation to make AI a company-wide core competency and competitive advantage. This report will evaluate the state of the enterprise MLOps platforms market, exploiting current technology trends, provider approaches, and future requirements.

Omdia view

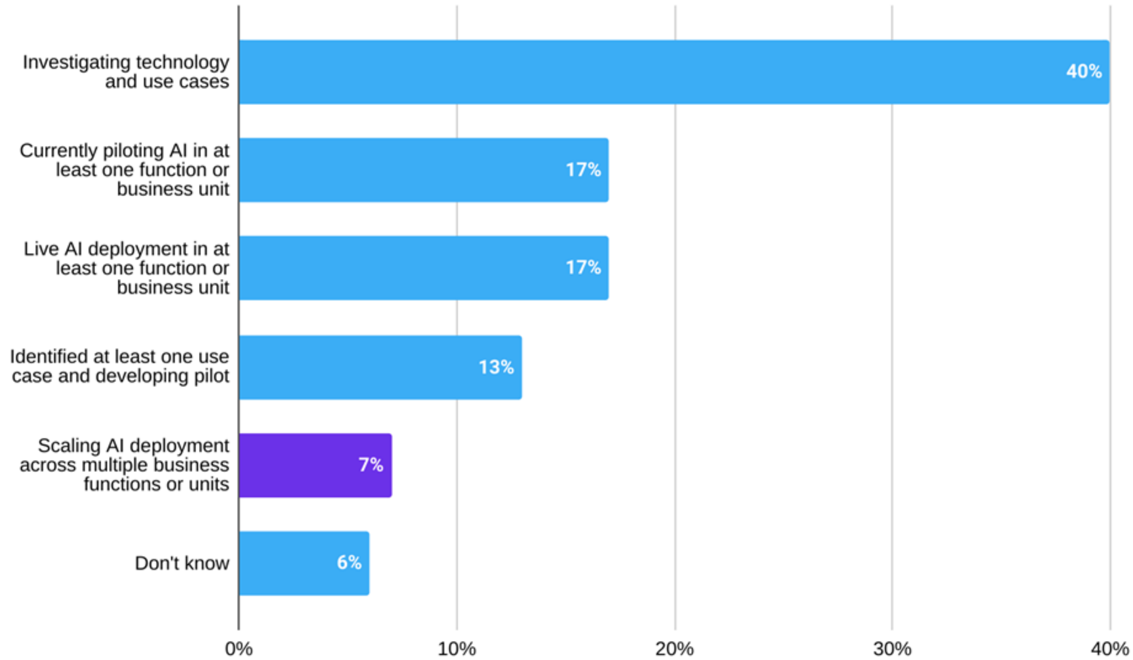
Whether building a simple quarterly sales forecast or automating a highly performant fraud detection system, enterprise AI practitioners have at their fingertips an abundance of technology to simplify, streamline, and even automate complicated data science workflows beginning with data preparation and ending with predictive model training and tuning. Thanks to public cloud platforms such as Microsoft Azure, Google Cloud Platform, and Amazon Web Services (AWS), in just a few minutes companies can provision all the supporting storage, processing, and ML resources necessary to begin building AI outcomes. And with technological innovations such as AutoML, which can automate feature engineering, model selection, model tuning, and model training processes, companies with minimal data science experience can quickly begin to create impactful AI projects.

This has led to broad AI exploration within the enterprise, as reflected in a recent Omdia online survey (*Omdia AI Market Maturity, 2020*) of 365 enterprises, which found that 70% of enterprise AI practitioners were exploring AI use cases, had identified at least one use case to pursue, or were piloting at least one project. However, the same study revealed a significant problem. Despite this acute willingness to explore AI, only 7% of those surveyed reported scaling live AI implementations across business functions or business units (see Figure 1).

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Figure 1: AI adoption in the enterprise

Figure 1: AI adoption in the enterprise1



n = 161

Omdia AI Market Maturity, 2020

End-user companies were asked, “Q: What is the state of AI deployment at your company today?”

© 2020 Omdia

Why is this? Bringing an ML project to production and keeping it there demands far more than readily available data storage, AI hardware acceleration, and some automated data science routines. Successful enterprise ML at scale demands the careful orchestration of a complex tapestry made up of people, processes, and platforms, an effort that does not end when an ML solution goes live but instead continues for the life of the solution. Omdia believes that deploying and maintaining ML solutions at scale over time demands the operationalization of the ML lifecycle through the adoption of a rapidly emerging class of enterprise MLOps platforms that combine DevOps technologies and practices within a highly centralized and collaborative environment that covers the entire ML lifecycle and all of the user roles necessary to create the ML solutions.

Without investment in an enterprise MLOps platform, Omdia believes that enterprise practitioners may find themselves falling behind better-prepared rivals and unable to leverage AI as a means of optimizing and innovating across both business and operational concerns. Worse, companies that are not able to use AI systemically as a means of optimizing processes, reducing cost, and identifying new opportunities, might fail to adapt to unanticipated and unprecedented market disruptions such as the COVID-19 pandemic.

Continuing coverage of this important and rapidly evolving marketplace, in the fourth quarter 2020, Omdia will publish a comprehensive, competitive evaluation of enterprise MLOPs platforms in the fourth quarter 2020, assessing how solutions from technology providers operationalize the entire ML workflow and

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.

lifecycle, including ideation and design, reparation and exploration, development and experimentation, validation and deployment, monitoring and revision, as well as management and governance.

Key messages

- Deploying ML at scale in the enterprise is a multi-faceted endeavor covering people, process, and platform concerns.
- DevOps practices and technologies show promise in solving many ML operational concerns such as project deployment, testing, and monitoring.
- Enterprise MLOps platform can successfully apply DevOps principles to the task of operationalizing ML, despite numerous ML operational, collaborative, and infrastructure complexities.
- The enterprise MLOps platform marketplace is growing rapidly, but solutions are still evolving to tackle important standardization, collaboration, and integration challenges.

Recommendations for buyers

- **Cloud-native computing**: Enterprise AI practitioners, in evaluating MLOps platforms, should invest in solutions that are fully cloud-native both in how they are built and how they are used to build ML outcomes. For example, solutions should leverage containerization, microservices, automation, and API interface and should run on all public cloud platforms and private cloud container platforms such as IBM Red Hat OpenShift. The use of cloud-native architectures will ensure compatibility with external, supportive solutions such as data security, quality, and governance tools. It will even enable buyers to more readily combine best-of-breed services found in multiple MLOps solutions. It will also enable practitioners to more readily use a wide array of supportive infrastructure resources such as AI acceleration hardware, and premises, edge, and cloud deployment platforms.
- **AutoML**: No longer limited to academics looking to build more efficient deep learning (DL) neural networks, AutoML can now help business subject matter experts take responsibility and ownership for many data science demands such as feature engineering and model training. Many companies see this kind of data science democratization as an effective means of bringing the business into the MLOps lifecycle not only as an adviser but also as an active stakeholder.
- **Keep an open mind:** Enterprise buyers looking to operationalize ML at scale should carefully consider three unique product avenues. First, if they are well-versed in cloud-native and data science technologies, they can self-assemble an MLOps solution using tools such as Anaconda, MLFlow, or DVC. Second, they can make use of the work done by large enterprises and adopt productized MLOps platforms such as Netflix Iguazio or Uber Michelangelo. Third, companies can take a more direct approach and opt for a pure-play offering from vendors including Databricks, H2O.ai, DataRobot, Cnvrg.io, Cloudera, SAS, Dataiku, and many more. Regardless, buyers must bear in mind that MLOps platforms are not now nor will they work as a simple, plug-and-play piece of software. Broaderintegration, customization, and augmentation requirements will be necessary to put any MLOps solution into practice.
- **Responsible AI:** Although many enterprise MLOps platforms are actively building in mechanisms specific to the creation of safe, secure, reliable, transparent, and fair ML outcomes, it's still early days

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.

with a great deal of work outstanding. In the meantime, buyers should consider creating an AI center of excellence (CoE) that directly addresses responsible AI, with clear lines of authority, following the European Union's oval guidance on the matter: "Develop, deploy, and use AI systems in a way that adheres to the ethical principles of: respect for human autonomy, prevention of harm, fairness and explicability. Acknowledge and address the potential tensions between these principles."

Technology overview

The challenge of building successful AI outcomes in the enterprise

Why is there such a sizable gap between isolated experimentation and widespread implementation? At a very fundamental level, AI practitioners must run a complex gauntlet of challenges in creating even the most mundane ML task. Whether training a pre-built chatbot model for customer service or coding a fraud detection DL algorithm from scratch, at each step in the ML lifecycle, a single misstep could slow development, elevate cost, or render the final outcome moot. Worse, a single misstep could deliver misleading or erroneous results that might go undetected even as the solution enters production. Above basic processes, a host of challenges and potential pitfalls await enterprise AI practitioners, including use case selection and performance measurement, project ownership, skills acquisition, and privacy/security compliance. It is however the mundane, day-to-day workings of ML that pose the greatest, ongoing challenge for companies that want to become data- and AI-driven.

Numerous missteps confront enterprise AI practitioners looking to build ML outcomes. Consider an extremely simple ML project that aims to predict which customers are least likely to renew an annual service subscription using historic customer data. For this example, there are at the most fundamental level three important phases of development:

- Data acquisition and preparation
- Data science exploration and experimentation
- Predictive model deployment and monitoring

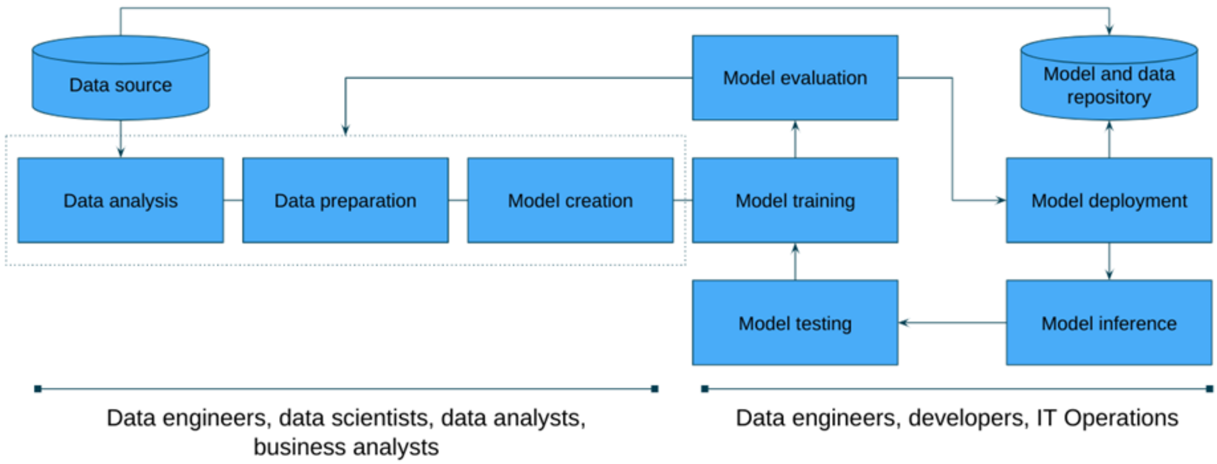
With this simplified customer churn example, an AI practitioner, usually a data engineer, must obtain, clean, transform, and label the necessary data for training and testing the predictive model. A data scientist must then explore the data, further engineering data features, and then embark on a series of iterative experiments to find and optimize the most applicable ML model that will yield the desired outcome. For this example, once a model reaches the desired accuracy, the data scientist can publish their findings, helping marketers best decide how to communicate with at-risk customers.

This model, however, could just as easily run continuously in production, using live data to deliver an ongoing view into the state of the subscriber base. In this situation, provisions must be made to host the running model, deliver predictions programmatically, and monitor the model for accuracy over time (see Figure 2).

Figure 2: Enterprise ML workflow

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Figure 2: Enterprise ML workflow2



© 2020 Omdia

With these very fundamental complexities, the relatively low penetration rate of AI as a company-wide endeavor begins to make perfect sense. The challenge for companies looking to move beyond the realm of basic AI proof of concepts (PoCs) and pilot programs rests within three key concepts for any given ML experiment or AI implementation:repeatability, scalability, and surety.

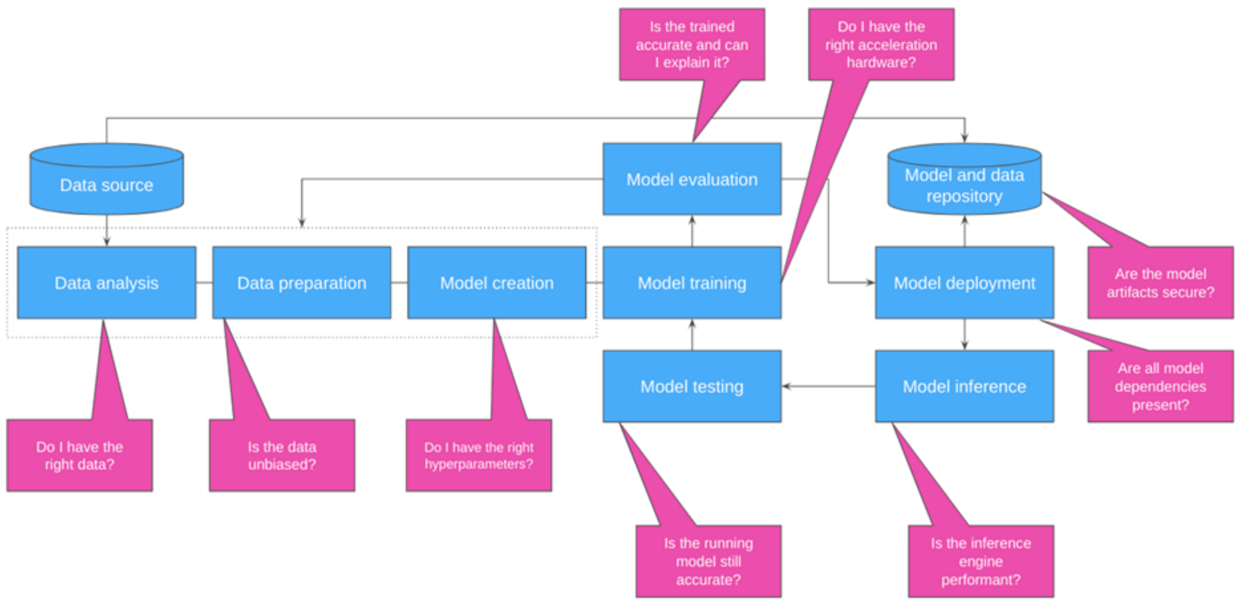
- Repeatability: It is difficult for AI practitioners to repeat what is a highly investigational methodology filled with stops and starts, dead ends, and unforeseen avenues of exploration. Unmanaged code, often written in Python and R, lives ungoverned within various Jupyter notebook implementations, making it nearly impossible for anyone but the original author to track and manage the code over time. Add to this the challenge of maintaining versioning across a wide array of libraries and frameworks, which change from project to project, and it’s easy to see how lessons learned in one project do not readily carry over into future endeavors.
- Scalability: It can be notoriously difficult for companies to effectively manage resource requirements, such as AI acceleration hardware for both development (training) and deployment (inference), because these tasks are themselves dependent on a myriad of malleable conditions. The same holds true for all supportive storage and processing resources, such as database instances, data pipeline processing, and inference engine execution. The high degree of entanglement makes it difficult for IT managers and CTOs to predict and therefore manage costs, a difficulty that grows exponentially as new AI projects enter development and production.
- Surety: It can be difficult to trust AI business outcomes altogether, owing to a lack of transparency within many DL predictive models, unchecked biases lurking in both data and model alike, poor code documentation across the project lifecycle, and inadequate testing of models prior to deployment. Even if an organization successfully tackles these and other similar challenges during deployment, maintaining a level of confidence over time demands a high degree of vigilance, monitoring models to ensure that their efficacy does not diminish owing to changes in the supporting data or surrounding systems. For highly regulated industries, this kind of monitoring can demand the actual replication of a model’s output at a given time. This can be an impossible task for organizations that are unable to fully document the entire ML lifecycle, including data, data preparation, feature engineering, model selection, parameter tuning, and model testing.

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Unfortunately, there is a wide range of obstacles preventing even the most experienced IT organization from putting these concepts together in harmony. The notion of data science itself stands as an impediment, owing to its highly experimental and iterative nature. For each general phase in the AI implementation lifecycle, practitioners must face several challenges, answering many difficult questions (see Figure 3).

Figure 3: Challenging questions must be answered at each step in the ML lifecycle

Figure 3: Challenging questions must be answered at each step in the ML lifecycle3



© 2020 Omdia

This means that each AI project operates as a complex system that can be very difficult to predict or control. Any attempt to scale beyond a handful of AI projects only exacerbates these difficulties. It is interesting to note that software development found itself in a similar situation in the 1990s, when success depended heavily on navigating a complex and often unpredictable landscape of underlying hardware and software systems. Fortunately, the software development market evolved to incorporate highly flexible and manageable innovations such as agile development methodologies, collaborative version control practices, as well as software virtualization and containerization. These innovations, collectively described as operationalized development (DevOps), have evolved over the past decade to enable developers to greatly improve software development, deployment, and maintenance outcomes.

Operationalizing AI: moving from experimentation to implementation

Today, enterprise AI practitioners suffer from the same challenges that software developers faced back in the 1990s. They must tolerate a high degree of system- and organization-level complexity, and they must do so without any reliable means of controlling project dependencies at scale across the enterprise.

MLOps principles and practices

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Fortunately, many of the same DevOps techniques and technologies that modern enterprise software developers have come to rely on are available to AI practitioners. DevOps seeks to operationalize the software lifecycle through the application of several core principles:

- Architect for end-to-end accountability
- Automate everything that can be automated
- Start with the end solution in mind
- Always keep the customer front and center
- Support cross-functional, independent contributors
- Build to embrace continuous improvement

Collectively, these DevOps principles make up the spiritual core of MLOps, but each in turn points to just one singular starting point: a rigorously enforced means of managing all the metadata that feeds into and is created from within the lifecycle of a given ML project. Once established, this metadata can enable the adoption of numerous DevOps toolchain technologies including

- Collaborative source-code management and versioning system
- Centralized, managed artifact repository
- Model testing, deployment, and retraining platform
- Monitoring service for running inference engines
- Governance service to track, explain and replicate model outcomes

MLOps is not however DevOps. Unlike software code, ML models degrade over time, requiring continuous monitoring. For DevOps, an entire project can be versioned as a single unit. With MLOps, unfortunately, practitioners must version code, predictive models, and supporting data that changes over time. The idea of continuous testing in DevOps does not therefore map well to MLOps, because MLOps workflows demand continuous training and validation, two very different mechanisms.

MLOps: Applying DevOps to data science

Because of these differences, mapping DevOps principles and toolchain technologies to MLOps requires a complete makeover of established data science departments and a forced relationship upgrade between data scientists and IT staff members. Fortunately, because the AI marketplace was built on and continues to run on open source software, the barrier to entry for MLOps is surprisingly low. Moreover, MLOps does not demand a top-down, systemic approach and instead users can readily begin their MLOps journey by adopting a few simple tools.

To begin controlling Python and R resources, customers can put Microsoft's freely available GitHub platform to work with very little effort. GitHub enables data scientists to package their Jupyter notebooks in a format that can be versioned and shared collaboratively. To better manage library and language interdependencies within a Jupyter notebook, users can adopt Anaconda and to begin defining repeatable data pipelines that feed into and support Jupyter notebooks for training and inference, customers can employ DVC. Together these three freely available solutions can address MLOps concerns, such as repeatability, and can therefore help enterprises scale data science experience on a project-by-project basis.

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Using GitHub for code, DVC for data, and Anaconda for libraries will speed time to market, lower development costs, and mitigate many risks. Moving individual projects to production and growing beyond individual projects with MLOps, however, demands a more centralized architecture capable of applying the same ideas at scale over time. Of course, these and other open source projects can be combined and built on in a bespoke way to accommodate the complete MLOps lifecycle. However, this kind of undertaking lies well beyond the capacity (existing resources, skills, and experience) of most enterprises that stand on the cusp between pilot experimentation and critical use case implementation.

To close this gap, a large portion of enterprise AI practitioners are increasingly turning to lifecycle-complete MLOps platforms that emphasize repeatability, scalability, and surety. Anaconda, as an early market entry focused on managing data science environments, now tackles far more than the management of software packages through containerization, evolving into an end-to-end MLOps platform spanning development, collaboration, deployment, and governance. More recently, vendors, particularly those with experience in delivering supportive infrastructure, have gone a step further, investing in a combined MLOps and cloud-native workload orchestration layer as a means of driving digital transformation. This is the case with data center hardware and software vendors such as HPE, VMware, and IBM Red Hat that are seeking to operationalize ML solutions built in MLOps frameworks together with a data fabric, container platform, a management and security layer, edge services, and even services specific to the operationalization of IT itself (AIOps).

These innovations have led to a highly dynamic market that now holds many solutions that have built on open source projects such as Anaconda. This includes hyperscale cloud platform providers, Microsoft, Amazon, Google, and IBM. These vendors have been instrumental in maturing ML in the enterprise, driving the creation of many of the open source innovations that power many MLOps solutions in the enterprise, including TensorFlow and Kubeflow (Google), open neural network exchange (ONNX) and ML.NET (Microsoft), and AutoGluon (Amazon). These vendors have blended such innovations into a broader tapestry of public cloud services spanning database, transaction, and analytics, underpinned by an array of AI acceleration hardware services. As a result, many companies begin their MLOps journey aboard one or more hyperscale cloud platforms.

Interestingly the MLOps market also includes a small but influential group of enterprise AI practitioners that initially built their own MLOps solutions in order to succeed with AI at scale. These include Iguazio and Metaflow from Netflix, Michelangelo from Uber, and Flyte from Lyft. These coupled with the numerous pure-play offerings on the market from Databricks, H2O.ai, Datarobot, Cnvrg.io, Cloudera, SAS, Dataiku, and many more make up what is an exceptionally rich set of potential options for enterprise AI practitioners looking to purchase a ready-made enterprise MLOps platform.

Enterprise MLOps platforms: Key features and requirements

Steeped in the philosophy and practices of DevOps, these enterprise MLOps platforms take the operational benefits found in spot tools, such as DVC (data versioning), GitHub (code versioning), and MLFlow (deployment orchestration), and centralize these capabilities by tying them into a single set of project repositories (code, data, features, models, and other artifacts). This enables data engineers, data scientists, business analysts, developers, and IT professionals to work collaboratively while still focusing on their respective ML project tasks. This allows companies to bridge the gap between the experimental aspects of data science and the operational nature of software deployment, creating a living lifecycle (see Figure 4) capable of not only delivering continuous integration and delivery, but also expanding to include integration, training, delivery, and monitoring.

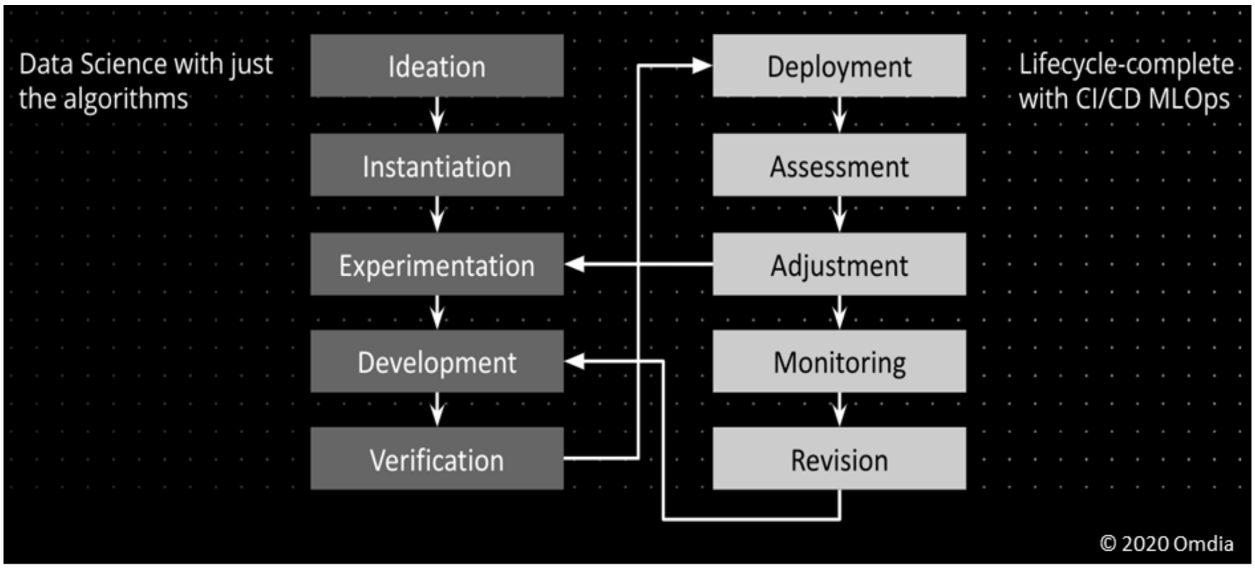
- Integration: Evaluating and validating all UML artifacts (data, models, etc.)

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.

- Training: Training during development and retraining after deployment
- Delivery: Packaging all artifacts along with supportive infrastructure (storage, compute, environment, etc.)
- Monitoring: measuring all ML artifacts in production, making ongoing adjustments, fulfilling compliance requests, etc.

Figure 4: MLOps lifecycle

Figure 4: MLOps lifecycle4



The task of evaluating a comprehensive solution such as an enterprise MLOps platform therefore requires AI practitioners to contend with a sizable number of features and capabilities that can vary considerably from solution to solution. Most enterprise MLOps platforms do, however, share several common capabilities (see Figure 5).

Figure 5: Enterprise MLOps platform key characteristics

Figure 5: Enterprise MLOps platform key characteristics5

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.

- Incorporate or fully integrate a central repository for the management of ML models, features, algorithms, code, as well as other project artifacts and metadata.
- Provide a central data repository (internal or integrated) as a means of facilitating data security, governance, auditing, and re-use. It is a highly desirable platform trait to have the flexibility to integrate with existing customer investments.
- Furnish facilities to provision, orchestrate, and manage system resources, both software and hardware (e.g. AI acceleration hardware) across the full ML lifecycle.
- Utilize a go-to-market strategy that emphasizes a subscription-orientated utility pricing model and dedication to cloud-native, hybrid architectures for deployment, development, and integration. This includes the ability to run on major cloud providers, on-premises, and hybrid scenarios.
- Employ a substantial degree of proprietary technology, while still leveraging open source software (development tooling, container software, ML frameworks, etc.) where advantageous.
- Support ML project development across a wide range of horizontal use cases. This often includes the productization of horizontal and vertical use cases leveraging pre-built resources and professional services.
- Accommodate disparate model deployment scenarios, spanning server, edge, and device footprints.
- Provide an ML quality, security, and performance evaluation and monitoring capability, tailored to particular problem types with pre-built functionality.
- Offer professional services engagements but not rely exclusively on those services to deliver product.

© 2020 Omdia

Further MLOps considerations and concerns

Architectural complexities

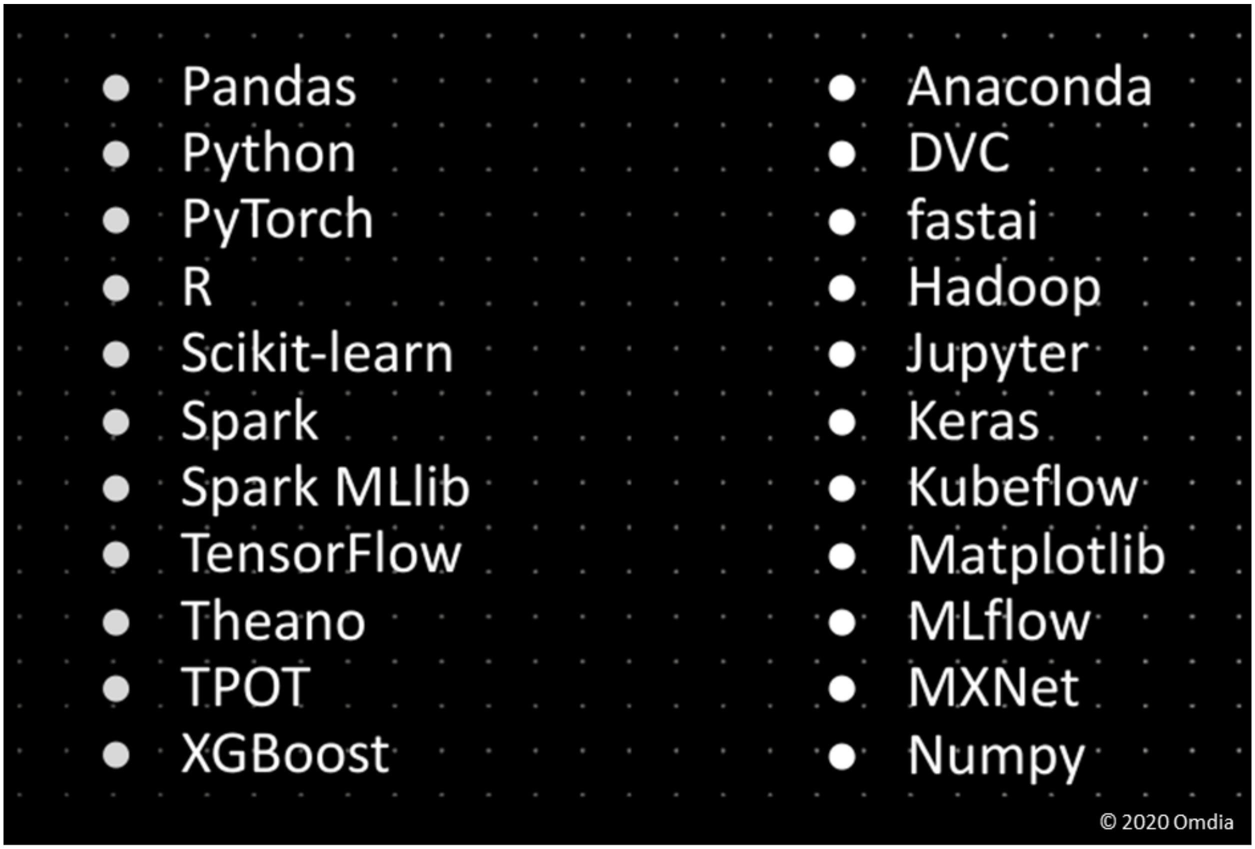
From an infrastructure perspective, enterprise MLOps platforms present practitioners with a complex tapestry of interwoven and interdependent elements. There are numerous subscription and purchasing structures used by services firms, platform vendors, hardware providers, and other market participants. Even with a growing number of lifecycle-complete AI development platforms available, most enterprise AI practitioners must contend with a wide array of providers and supporting technologies. All these solutions provide enterprise practitioners with an operationalized framework upon which to build AI outcomes. For example, while some integration and customization service work might be included within an AI solution, an additional, costly professional services component could be required to ensure a smooth delivery. In fact, many MLOps providers incorporate professional and managed hosting services into their core software subscription contracts as a means of ensuring that customers realize a marked return on investment.

AI implementations, successful or otherwise, are powered at least in part by open source software. Data scientists and data engineers actively use a staggeringly complicated collection of libraries, frameworks, and tools. Each open source technology (and numerous versions of each technology) must be provisioned and supported by enterprise IT practitioners on a project-by-project basis, creating an uneven patchwork of support contracts and version dependencies that must be managed (see Figure 6).

Figure 6: Important open source ML and MLOps projects

Figure 6: Important open source ML and MLOps projects6

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.



To combat these complexities, enterprise practitioners are increasingly turning toward MLOps platforms that can demonstrate a significant level of dedication to open source. This dedication often comes through direct partnerships between the cloud provider and vendors built on open source offerings such as Confluent (Kafka), DataStax (Cassandra), and Elastic (Elasticsearch). Cloud providers will often contribute to and offer their own implementations of popular open source projects. They will also often allow for direct support of these projects within their own proprietary offerings, as is the case with both Microsoft and Google that support Keras within their own DL frameworks.

This kind of openness also extends to computing infrastructure. Recognizing the importance of matching an AI project with the correct AI acceleration hardware regardless, cloud platform providers have been quick to invest in a wide array of chip architectures both their own and those offered directly by chip manufacturers such as Nvidia, Microsoft, Google, Xilinx, and Intel. AI practitioners building on top of cloud platforms such as AWS and Microsoft Azure can orchestrate containerized training and inference workloads that rely on AI-specific hardware spanning GPU, CPU, FPGA, NPU, and other chip formats.

Organizational concerns

Even with a robust enterprise MLOps platform at the ready, organizationally companies must weave together a similarly complex pattern of human expertise, including software developers, IT architects, data engineers, data analysts, data scientists, business analysts, security specialists, and domain experts. With the application of AI itself, as with AutoML, some of these tasks associated with each role can be augmented, simplified, and (in some limited instances) fully automated. Still, to put a proper AI team in place and to enable that team to operate across the business at scale demands a high degree of cross-departmental coordination and ongoing oversight.

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.

The ownership of AI within an organization is key to the way in which AI projects are implemented and managed on a day-to-day basis. Centralized management structures for AI require a chief AI officer in charge of AI projects, spearheading the management and implementation of AI within an organization. This role, however, is not yet a common practice. Instead, it is often filled by a chief data officer or chief technology officer. Centralized AI management allows for a “paintbrush” approach to AI where AI can be applied to multiple silos within an organization, but central management risks not clearly understanding the key business metrics or needs within each of the silos. Centralized management is, however, a good approach to attracting the best talent, with AI engineers looking to join a company where the “head of AI” is someone that has a major standing in the AI community.

For the most part, partially decentralized AI is a better approach to follow because it retains central authority but grants departmental autonomy. This hybrid management model gives departments the freedom to define and drive their own AI strategy rather than have central management dictate terms. Also, in the long run, as AI becomes the default way of running and managing a business, the CEO can serve as the default chief AI officer or head of AI, and therefore create a separate business function for AI in a centralized approach separate from the CEO.

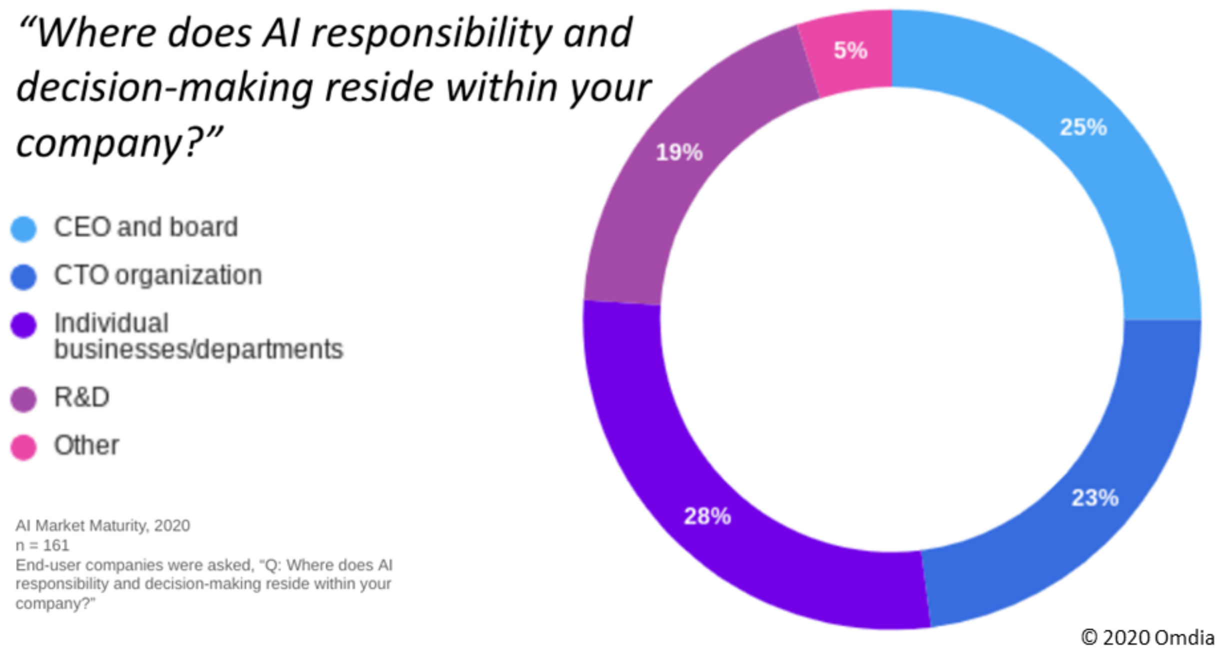
Until that time, Omdia recommends that practitioners establish an AI center of excellence (CoE), staffed with a core set of data scientists, data engineers, and other domain experts, that falls under the auspices of a chief AI officer. This group can oversee without dictating total control over individual departments or endeavors. This encourages business units to employ their own domain experts as needs demand.

According to Omdia’s AI Market Maturity, 2020 survey of enterprise AI practitioners, this coordination is still difficult to come by. Today, there is no absolute consensus on where AI responsibility resides within end-user respondents, because decision-making appears to be evenly distributed across different power centers (see Figure 7).

Figure 7: AI decision-making

Figure 7: AI decision-making7

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.



It is likely that this decision making will consolidate over time, perhaps moving completely out of R&D as many companies begin to think about AI expertise and intellectual property (IP) as a core competency. Still, for most organizations, AI will remain diffused across various business stakeholders with overall control passing back to some central CoE or other guiding authority.

What's next for MLOps?

As the MLOps platform marketplace matures, and as more companies gain experience in deploying AI at scale across numerous lines of business, these organizational best practices will become clearer. For example, many companies are plying technologies such as AutoML to democratize data science, putting more control into the hands of domain experts who are best situated to understand the nuances of the underlying data and needs of the organization. These practices will dramatically improve the value of an enterprise MLOps platform. Even so, these platforms have plenty of room to grow.

First and foremost, the MLOps marketplace overall is lacking in standardization. Standards capable of bringing together disparate, proprietary software are lacking. Similarly, standards covering ML models and supportive artifacts are only now starting to find their way into the marketplace. Early but restrictive efforts such as predictive model markup language (PMML) have fostered subsequent efforts such as portable format for analytics (PFA) and ONNX. However, a substantive consensus among ML technology providers has yet to emerge, leading to proprietary integration efforts such as ModelOps' agnostic scoring engine FastScore.

Furthermore, MLOps vendors are still heavily reliant on external solutions for well-established data requirements such as data governance, management, and security. Vendors are, however, investing in their own capabilities to address data and ML quality, transparency, understanding, and fairness concerns, plying AI itself to tackle problems such as model explainability. But even here, coverage remains inconsistent from solution to solution and in some cases, such as with identifying model or data bias, only targets select algorithms or use cases.

And lastly, while most MLOps solutions provide for collaboration among disparate roles, collaboration efforts have yet to properly embrace advanced requirements such as actionable workflows, artifact

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.

certification, and project documentation. Regardless, enterprise MLOps platforms can greatly reduce the burden that enterprise practitioners must carry in order to successfully create ML projects. They can also help companies better manage (even reduce) the level of investment in both the expertise and resources necessary to support the models in production. The result? Companies with little data science expertise can more readily get started with ML, and companies struggling to evolve beyond isolated projects can more readily scale beyond the limits of experimentation to make AI a company-wide core competency and competitive advantage.

Download MLOPS-20-Omdia - 864.53kb | pdf

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Citation policy

Request external citation and usage of Omdia research and data via citations@omdia.com.

Omdia consulting

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Omdia's consulting team may be able to help you. For more information about Omdia's consulting capabilities, please contact us directly at consulting@omdia.com.

Copyright notice and disclaimer

The Omdia research, data and information referenced herein (the "Omdia Materials") are the copyrighted property of Informa Tech and its subsidiaries or affiliates (together "Informa Tech") and represent data, research, opinions or viewpoints published by Informa Tech, and are not representations of fact.

The Omdia Materials reflect information and opinions from the original publication date and not from the date of this document. The information and opinions expressed in the Omdia Materials are subject to change without notice and Informa Tech does not have any duty or responsibility to update the Omdia Materials or this publication as a result.

Omdia Materials are delivered on an "as-is" and "as-available" basis. No representation or warranty, express or implied, is made as to the fairness, accuracy, completeness or correctness of the information, opinions and conclusions contained in Omdia Materials.

To the maximum extent permitted by law, Informa Tech and its affiliates, officers, directors, employees and agents, disclaim any liability (including, without limitation, any liability arising from fault or negligence) as to the accuracy or completeness or use of the Omdia Materials. Informa Tech will not, under any circumstance whatsoever, be liable for any trading, investment, commercial or other decisions based on or made in reliance of the Omdia Materials.

CONTACT US

omdia.com

askananalyst@omdia.com

© 2020 Omdia. All rights reserved. Unauthorized reproduction prohibited.