MLOps Software Buyer's Guide

Contents

What is MLOps, why is it needed, and why now?

MLOps or Machine Learning Operations is a set of processes and practices that automate and scale the deployment and management of machine learning models in production environments. By bringing DevOps principles to machine learning, it enables a faster development cycle, better quality control, and the ability to respond to changing business requirements.

Moreover, MLOps is much more than model design and development. It also includes data management, model retraining, monitoring of the model and continuous development.

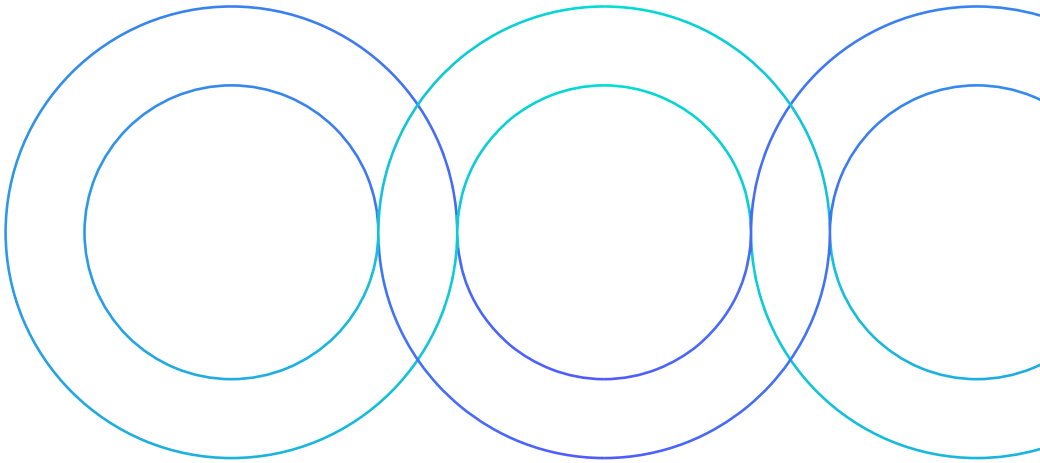
The origins of MLOps goes back to 2015 from the research paper "Hidden Technical Debt in Machine Learning Systems" and since then, there has been no looking back. What began as a broad set of practices has, over time, evolved into an independent approach to machine learning lifecycle management.

As per a 2019 report from MIT, 7 out of 10 executives whose companies have made investments in AI reported a minimal impact of the technology on their business. Why? Because developing machine learning models and putting them into production environments are two completely different things. The cost of poor deployment can lead to risk, not being able to scale, and delays to projects taking months and negatively impacting organisations' top and bottom lines.

Data scientists can create the most effective models for machine learning problems but since they are not dedicated developers, they have limited knowledge of the tools and skills to test, deploy or maintain models. This is where MLOps plays a crucial role.

It facilitates smooth communication and collaboration between operations professionals and data scientists. With MLOps software, it also becomes easier to align models with business and regulatory requirements.

While a few companies have succeeded in generating value with AI, most companies have a hard time with it. MLOps helps companies unlock potential, manage risks, and minimise the bottlenecks associated with machine learning.



Bottlenecks limiting organisations' machine learning

Siloed teams and tools

MLOps software improves the lifecycle of the ML program by automating the pipeline process and reducing the time to production. It's a process in itself with a goal to create continuous development, deployment and improvement of machine learning models.

However, when there are multiple teams and tools or activities, a project or process can become blocked or slowed down in each silo. Various tools are often built for specific tasks and teams, but when a process involves multiple tools that don't communicate with each other or teams that aren't working in a transparent way, this is when projects start to become misaligned and projects delayed.

Even within teams, using multiple tools makes it difficult to get all the information in one place. Each set of information is stuck in its own tool silo. For instance, some tools focus mainly on the deployment of the models, while others on training.

MLOps software helps break down the walls of silos by uniting multiple teams in a single framework. The main objective is to ensure that everyone involved in an MLOps strategy is aligned on the who does what, technology and processes, and has the visibility to collaborate both effectively and efficiently across the machine learning lifecycle.

Getting trained models into production

Every machine learning model is designed with respect to various machine learning platforms, programming languages, and environments.

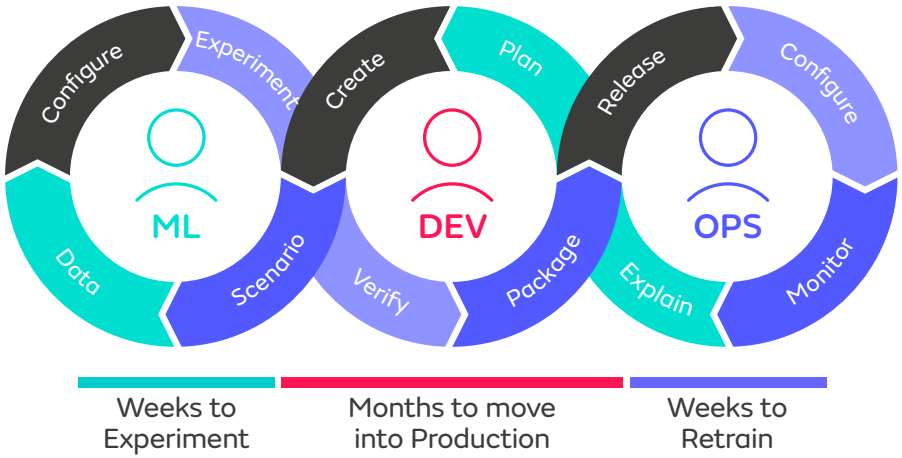
Many organisations get stuck at the prototype stage and are not able to serve this into production, or if they do, it can take months. Spiralling costs and skills shortages mean that despite an ever increasing number of companies using machine learning, just 1 in 10 machine learning models ever make it into production. It's only when a model is put into production that it starts to generate ROI and provides any value to the business.

One of the most common reasons for this challenge is the difference in skills between data scientists, who build and train the models, and engineers who develop and deploy production-ready ML applications. On top of this, it becomes even more difficult to deploy and manage machine learning infrastructures as the organisation evolves and grows.

This is where MLOPs can help by providing standardisation and visibility across ML infrastructure, and the ability to serve models in production at scale. Implementation of standard practices tailored to the ML infrastructure can also help reduce risk.

Steps of MLOps

Long deployment pipelines slowed by manual workflows can shrink revenue. It's important to understand that the MLOps process is ongoing with optimisations, and never necessarily complete. The seven steps of MLOps can be described in this diagram.



1. Definition of the business need
This is the origin of every machine learning project. It defines the objectives of the project, the success criteria and the teams required for the project like data scientists, data engineers, machine learning engineers, IT executives, etc.

2. Data preparation
Structured or unstructured, data is the root of any successful model. It is one of the key factors in any ML model.

3. Experiments
When it comes to MLOps frameworks, this is actually the first step. Data scientists test multiple models, multiple hyper parameters, multiple features.

4. Model validation
Model validation ensures that there will be no surprises once the model goes live.

5. Model deployment
If successful, model deployment will bring a ROI to the organisation in line with the stated objectives. The success of this step is totally dependent on how clearly you have specified the requirements and elements you need at the time of delivery.

6. Model monitoring
Monitoring verifies that input data, predictions made by the model and the output data matches with what is expected.

7. Model retraining
Model retraining is a necessity because model performances will usually decay over time which in turn leads to lower performance.

Unlocking business potential

Organisations are trying to efficiently build, serve and manage machine learning models throughout the machine learning lifecycle. Scalability, collaboration, deployment and automation, reproducibility of models, monitoring and management are a few of the benefits teams wish to realise from MLOps software while walking on the path to turn R&D into ROI. Depending upon company size, their MLOps infrastructure can be represented by something as simple as a set of vetted and maintained processes, and use cases will vary by industry.

Among the various positive aspects of MLOps software, the overarching benefits relate to the ability to serve models in production and scale sustainably. With this ability, businesses can accelerate their performance but also future-proof their operation. Other benefits include:

- Breaking down team and solution silos
- Self-service and control over the lifecycle
- Reducing speed to production (standardised method that can be reproduced)
- Democratisation of AI – taking models from an R&D team, making predictions useful to people beyond the data science teams. ML can be utilised by the entire business
- Having full data lineage and transparency – being able to track each model back to the data used to train it
- Saves the time required to build your own tools
- Optimised model performance and governance

Finding the right solution

There is no one size fits all approach when it comes to choosing the right MLOps software. There are internal and external factors that will influence the questions you ask and decisions you make, including company size, team skills, sophistication, and industry to name a few.

Open Source vs Enterprise

Open source is a term used to describe software for which the original source code is made freely available. The public is allowed to copy, modify and redistribute the source code without paying any royalty or fee. Enterprise software is software licensed under exclusive legal right of its owner. When purchased, the purchaser gets the right to use the software under certain conditions. However, the purchaser cannot modify or redistribute the same.

Both have their advantages and disadvantages but what works for an organisation depends upon the functionality it is looking upon and skills internally. In a recent survey, it was found that 78 percent of the companies run open source software which clearly indicates the widespread adoption of open source.

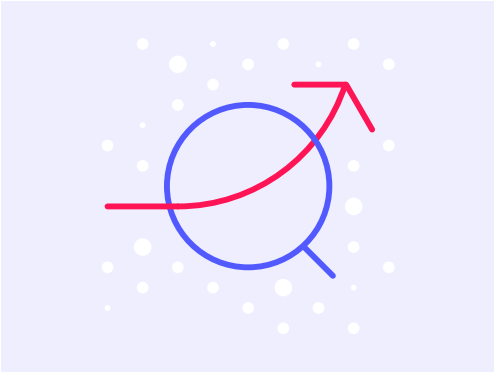
In some ways, open source outperforms enterprise software and the reason is being free and flexible. However, there are hidden costs most notably the cost of resources in building a platform internally and the risks associated with it breaking or not achieving what it set out to, especially with open source software not being standardised. Those evaluating open source who like to manage the risk and scale with assurance should consider Seldon Core Enterprise, which as well as SLAs, also offers service management to help you achieve your desired machine learning infrastructure and goals.

Enterprise software is ready built, standardised, often has more features but is sometimes less configurable. It is more powerful, quicker to use and usually updated regularly with new features to aid customer experience. Those looking for enterprise MLOps software with a user interface that makes it seamless to deploy, monitor and explain models should consider Seldon Deploy.

Maturity

MLOps is not only about technology, but also about the processes, operations and people involved. Using the maturity model will help you assess the current effectiveness of a team or tool to figure out what capabilities are needed to acquire next in order to improve performance, and become more mature as an organisation when it comes to machine learning.

Regardless of where you fall on the maturity model, you should be asking questions around your long term machine learning plans and objectives, and consider the bigger picture to future proof your infrastructure. Most organisations are at the serving maturity stage, but something we're seeing at Seldon is more interest in machine learning monitoring, and explainability features in highly regulated industries such as financial services.



Managing models as you scale

Model management is a part of MLOps. ML models should be consistent, and meet business objectives at scale. Thus, it is important to have an easy-to-follow policy for model management. ML model management helps in the development, access, versioning and deployment of ML models.

When data scientists work on ML models, or apply them to a new domain, they run countless experiments with different model architectures, data sets and parameters to get the most optimum model. Keeping track of such experiments is crucial because without them, it is impossible to compare and select the best solution.

Effective model management helps teams and organisations:

- Proactively address common business concerns
- Enable reproducible experiments
- Support reusability
- Store, manage and search for features efficiently
- Manage users based on access rights

Optimising models

Over time, models are likely to decay as the data they were trained on becomes less relevant. This is when models start to perform less effectively and are due to be retrained. Something that is now becoming more common, and where a lot of organisations have shifted to in their machine learning maturity is the monitoring stage.

Organisations that are deploying ML and are not monitoring drift are at best slow to respond when performance dips or changes unexpectedly. An important consideration would be alerting tools to notify and give developers a faster feedback loop so they can improve future iterations more quickly.

Another important consideration is visibility through dashboards and metrics to give confidence in the performance of models and understanding of how your ML is performing against objectives. Finally, being able to run deployment strategies such as canaries or AB testing, will help constantly optimise models to ensure that the best performing model is running.

Understanding models

The final stage of MLOPs maturity is around explainability, and is something that is prevalent in (although not limited to) highly regulated industries where compliance and transparency are critical.

MLOps software allows your stakeholders to understand why a model is performing the way it is. Having the right information at the right time allows you to answer to internal auditors and external regulators more quickly, and avoid non-compliance fines through successful model governance.

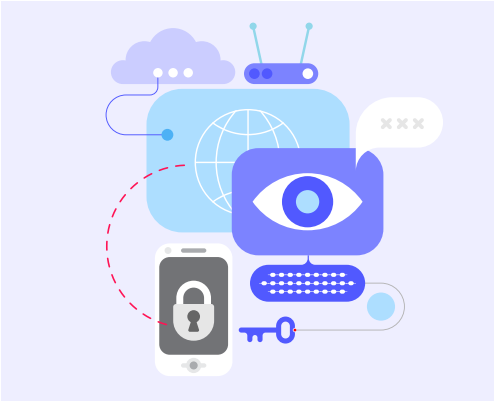
An important consideration here is to be able to understand model behavior and influences, and audit your model when it begins to become less effective – and to be able to do this quickly and easily. Another question you should consider is how you can ensure full accountability and traceability across all models.

As the number of models in production increases, this correlates as a challenge, and if a model stops performing or results in a compliance breach, it is important to be able to track this back, including version history.

Security

MLOps is all about accelerating the machine learning lifecycle in a secure and trusted way. MLOps software must provide organisations the confidence of being able to ensure that data is secure and not shared openly, user access controls are developed properly and teams aren't overlapping.

To get models to production faster in a secure way, it is important to leverage multi-factor authentication, role-based authorisation, data encryption, and other security and privacy best practices. When businesses run the majority of their digital infrastructure on the back of ML, it is of paramount importance that the system is secure.



Best of breed vs end-to-end

A huge topic in MLOps is whether to go down the path of best of breed or end-to-end platforms. There is no right or wrong answer to this, and the outcome will depend on company and team sizes, skills, company strategies and direction.

Often for smaller companies or teams, it makes sense to go with an out of the box platform with the ability to train models as well - this will mirror the team setup where the number of team members is small and manage the full pipeline. However, a common mistake made is going down this route when the machine learning operation is fastly expanding and becoming more sophisticated in approach.

Lengthy implementation periods make it difficult to replace poor-performing technology within the suite with more effective tools, limiting its value and return on investment, which is why we must consider the long-term goals and bigger picture.

For many organisations either now or in the future, you may need stronger serve, monitor, or explainability features, or you may plan to run complex inference pipelines. This is where best-of-breed platforms are more suitable. Flexibility is also an important consideration, in terms of language, connectivity with your wider digital ecosystem or servers in the cloud or on-premise.

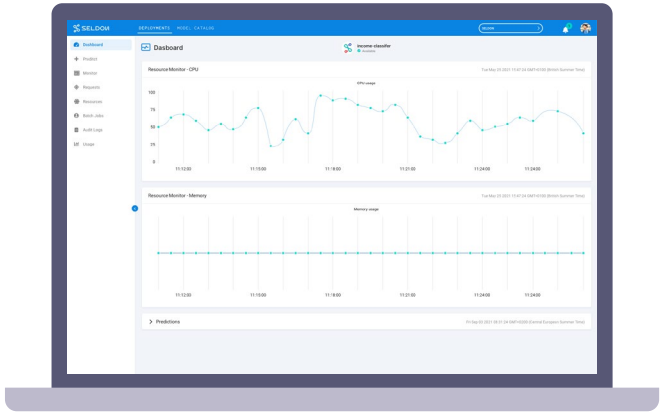
Unlock your business potential and manage risk with Seldon

Seldon moves machine learning from POC to production at scale. By enhancing time-to-value, your models can get to work up to 85% quicker. Seldon enables you to:

- Transform your organisation by managing your models and workflows at scale
- Boost productivity and reduce costs through improved model performance
- Automate workflows to improve customer experience and increase ROI
- Better understand changes to your data and reduce risk through monitoring and explainability

Those looking to scale their open source machine learning deployment with assurance can leverage Seldon Core Enterprise and benefit from SLAs, support and dedicated customer success to help accomplish your machine learning goals.

For organisations evaluating monitoring and explainability capabilities, and a ready-built platform, where these more advanced features can be powered by best-in-breed orchestration functionality, Seldon Deploy enables organisations to manage and monitor ML models to minimise risk, unlock business value, and scale efficiently.



Learn more

Connect with the Seldon team to learn how we can help your organisation grow strategically.

Email hello@seldon.io

Call +44 (20) 7193-6752

SELDON