

2021 MLOps Platforms Vendor Analysis Report



Executive guidance on selecting
the optimal technology solution
for supporting your MLOps journey

Intro

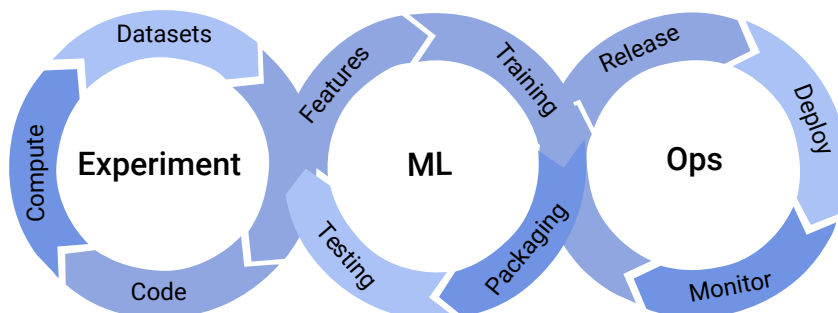
Defined as a separate discipline only recently, the MLOps movement has already moved from the conceptualization stage to corporate boardroom discussions on implementation. But while nearly every industry has now entered the AI race, adoption of operations best practices and thus competitiveness remains unequal.

AI and ML leaders today already have a better understanding of the MLOps lifecycle and the procedures and technology required for deploying new models into production (and subsequently scaling them). The State of AI 2021 report by McKinsey notes that 51% of AI high performers have standard tool frameworks and development processes in place for developing AI models.

However, laggards can still catch up and even overtake the leaders if they get strategic with process standardization and automation, which in fact is what most MLOps platforms promise to facilitate.

An MLOps platform is a centralized infrastructure hub featuring resources, tools, frameworks, and process automation scenarios for managing all steps of the machine learning lifecycle.

Machine Learning Lifecycle

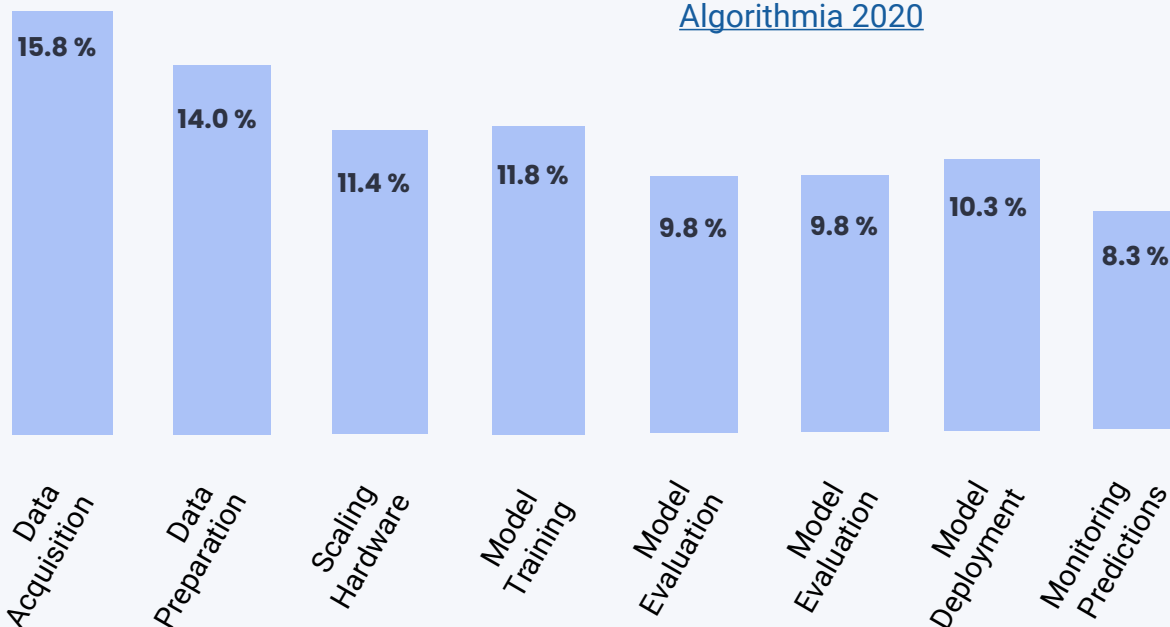


Back in the day, IDEs revolutionized the speed and quality of software development. MLOps vendors are now attempting to create a similar level of speed, convenience, and automation for data scientists with the ultimate promise of reducing model development and deployment timelines by 2X-10X.

Business leaders put high hopes on this promise: 73% believe that MLOps adoption will keep them competitive and 24% think it could propel them to the industry leader pedestal. In turn, this clear demand has attracted many new market entrants. The MLOps market is still young, but already has a host of established players (such as the major cloud vendors) as well as newer niche entrants.

In this report, we placed 10 MLOps vendors under a microscope to assess how they stack up in terms of:

- Infrastructure
- Data management
- Model development experience
- Model training and management
- Deployment capabilities



50%

Of AI experts use standard development tools and frameworks to create to AI models

[McKinsey 2020](#)

73%

Of business leaders believe that MLOps adoption will keep them competitive

[Forrester 2020](#)

\$4 bln

In annual revenue is projected to be generated by MLOps platforms by 2025

[Deloitte 2021](#)

1 to 30 days

42% of AI leaders require to deploy a new model

[Algorithmia 2020](#)

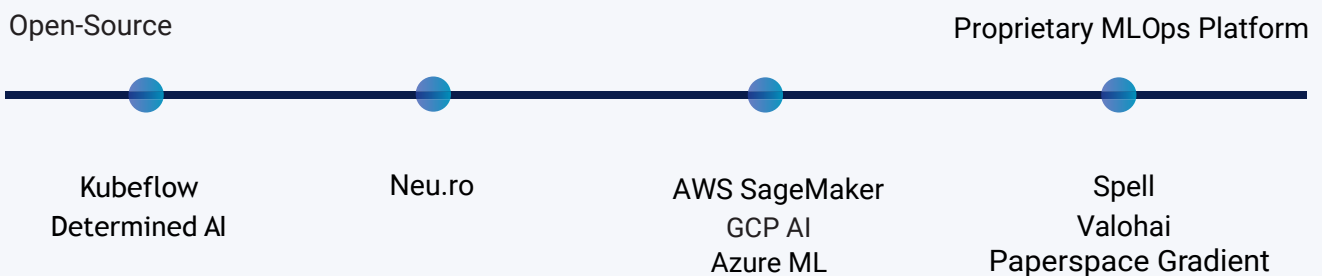
Open-Source vs Proprietary MLOps

Like other cloud products, MLOps platforms come with varying degrees of customization, extensibility, and native functionality. For our analysis, we ranked how each platform stacks in these terms.

By open-source, we understand platforms with fully available source code, distributed freely without any limitations on usage.

By proprietary, we categorized MLOps tools that are distributed as SaaS or licensed products, providing little-to-no access to source code and more limited customization options.

Open-Source vs Proprietary MLOps Platform



Kubeflow is an entirely free ML platform, offering machine learning pipelines and orchestration tools for running ML experiments on Kubernetes. Kubeflow also has tight integration with TensorFlow and other machine learning frameworks — PyTorch, Apache MXNet, MPI, XGBoost, and Chainer, among others.

The product can be nicely extended with different integrations and offers a high degree of freedom in customization. However, few MLOps features are available out-of-the-box apart from pre-made pipelines for ML workflows.

Limitations:

- Training jobs and model deployments need to be configured separately via TensorFlow
- No premade functionality for data ingestion and management

Determined AI has similar capabilities to Kubeflow. Distributed via GitHub, the platform features convenient hyperparameter search, intelligent job scheduling, and cluster sharing, and experiment tracking toolkit, log management, metrics visualization, reproducibility and dependency management. You can run Determined AI locally or in the cloud (compatible with all popular cloud services).

Limitations:

- It doesn't include native model serving and deployment functionality
- Model testing toolkit available
- Automated model training pipelines available
- No premade functionality for data ingestion and management

Neu.ro sits in-between fully open-source and proprietary MLOps platforms provided by established cloud vendors. The company has a unique value proposition and offers to custom-build and manage the MLOps for your business using a selection of open-source tools, and then hosting it on the infrastructure of your choice — locally or in the cloud. Neu.ro is fully cloud- and tool-agnostic, allowing you to assemble an ML development environment using a best-in-class set of tools. The proposed functionality, customizations, and support cover the end-to-end machine learning lifecycle from data management to containerized model deployments and model monitoring.

Limitations:

- **Neu.ro** distributes the core product (the orchestration layer) as an open core. It's free to use but only partially open-source. The end-to-end MLOps offering includes proprietary components.
- **Custom MLOps** platform setup and ongoing maintenance is a billable service. As well, customers may opt to pay for computing resources if they use the platform's infrastructure offerings.

AWS SageMaker, Google AI Platform, and Azure Machine Learning have similar PaaS offerings. While in each of these cases, you are bound to use the platform's cloud resources, these platforms do support a solid range of custom integrations with third-party tools, including open-source ones. The tradeoff, however, as with other open-source tools, is the lack of user-friendly business interfaces. All but Azure Machine Learning lack visual, drag-and-drop tools for citizen data scientists or collaborators.

Limitations:

- In all three cases, multi-cloud and hybrid deployment scenarios are limited and most users opt to run ML workloads and models on the platform.
- Deep product knowledge is required for each platform to effectively configure end-to-end MLOps capabilities.

Limitations:

- It doesn't support cloud providers or on-premises resources
- Data set uploads are limited to CLI
- Automated training pipelines not provided

Spell, Valohai, and Gradient (by PaperSpace) are fully proprietary platforms. They provide subscription-based access to a cloud MLOps platform and a set of native features such as collaborative workspaces, end-to-end solutions for running standard experiments, experiment tracking, and other features we discuss in-depth within the next sections. Amongst them, only Gradient can be self-hosted, while Spell, Valohai, and Floydhub installations are managed services.

Limitations of proprietary MLOps platforms:

- Limited customization/extensibility beyond the native features
- Limited support of deployment scenarios (mostly to cloud-based installations)

Interoperability rankings: Low to High

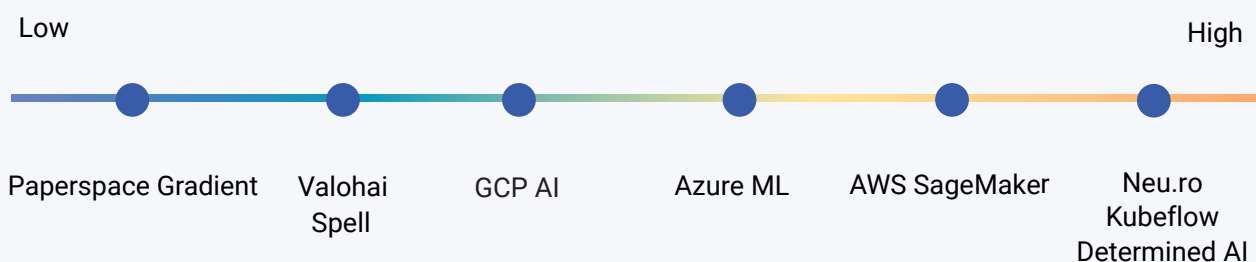
Like any young market, the MLOps platform landscape remains fragmented. There's a growing number of standalone open-source and proprietary solutions covering one leg or another of the MLOps lifecycle — such as data transformations, hyperparameter tuning, or model monitoring. These emerging mono-task solutions have their merit.

But we found that most data science teams look for well-integrated solutions. Thus, interoperability emerged as a key requirement for MLOps platforms.

Interoperability indicates the feasibility and ease of integrating various MLOps tools into a single consolidated setup for end-to-end operations.

Apart from ensuring greater flexibility and convenience for the teams, a greater degree of interoperability also offsets the operational risks. Choosing a platform that lets you incorporate different tools reduces the chances of vendor lock-in and enhances your ability to run more complex machine learning projects without the need to migrate to new infrastructure.

Interoperability: Low to High



Research Questions

In our research, we analyzed how well you can mesh together different MLOps stacks and which platforms have native or pre-built integrations with other popular solutions. Specifically, we asked the following research questions:

Does the platform have well-documented APIs, CLIs, SDKs?
Are those open-sourced?

Does the platform have native/productized integrations with other popular tools, or allow users to develop custom extensions?

Interoperability Rankings

Neu.ro, Kubeflow, and Determined AI scored the highest in terms of interoperability. That makes sense given the fact that all three solutions also top the open-source list.

Neu.ro has well-documented CLIs and an SDK. Both are open-source. However, the platform lacks a documented API. Still, this is compensated by a generous set of productized and documented integrations with popular MLOps solutions such as DVC, Pachyderm, MLflow, W&B, NNI, Seldon, and Algorithmia among others. Another 10 integrations are underway. Neu.ro is designed with interoperability in mind and strives to provide tools that make it easier to add new integrations.

Kubeflow, on the other hand, has an open-source and documented API + SDK. But CLI is still under development with the launch date yet to be announced. At present, the platform acts as an orchestration layer that anyone can augment with the toolkit of their choice. The majority of integration requests are driven by the community. Some of the currently available ones include Feast, Seldon, BentoML, Tekton (not to confuse with Tecton), MLflow, W&B, even Determined AI, and AWS Sagemaker — two other entrants on our list.


Determined AI supports the tech triumvirate of integrations — API, SDK, and CLI are open-source and well-documented. The number of pre-made integrations is slightly more modest than the earlier two entrants. These include DVC, Pachyderm, Data Lake, Algorithmia, Seldon, and Spark among others.

The main difference between the top 3 contenders is their product development vector. Neu.ro and Kubeflow have a stronger focus on delivering mature end-to-end machine learning orchestration, whereas Determine AI leans more towards acting as an “enabler” for building the baseline MLOps infrastructure and then extending it with the necessary third-party tools.

AWS SageMaker ranks a notch higher than other big-three cloud solutions. The platform has comprehensive APIs, CLIs, and SDKs for all mainstream programming languages. But none of these are open-source (understandably).

The platform has native functionality, covering all the steps of the ML project life cycle, plus several pre-made integrations with tools such as Tecton, MLflow, W&B, Seldon, and Algorithmia. Developers can further extend the platform with custom integrations.

Azure ML Platform has the bases covered when it comes to API, SDK, and CLI. Also, the company provides pre-made connectors for MLflow, DataBricks, GitHub Actions, and VS Code. Otherwise, Microsoft expects users to fully rely on the platform's native capabilities, which are rather extensive, as well as tools provided by the company's partners. Google AI Platform similarly has a well-documented API, SDK, and CLI. However, we didn't find any data about native connectors at the time of research.



The big three cloud companies have neat documentation for API, CLI, and SDK. However, most expect customers to fully rely on the platform's ecosystem of products without integrating third-party tools.

Valohai and Spell are both proprietary platforms with a closed product ecosystem. While Valohai has a documented CLI and partially documented API, primarily covering data injections. SDK does not seem to be available at the moment. Spell, on the other hand, has a documented SDK and CLI but lacks an API. Between the two, the platforms support only 4 integrations.

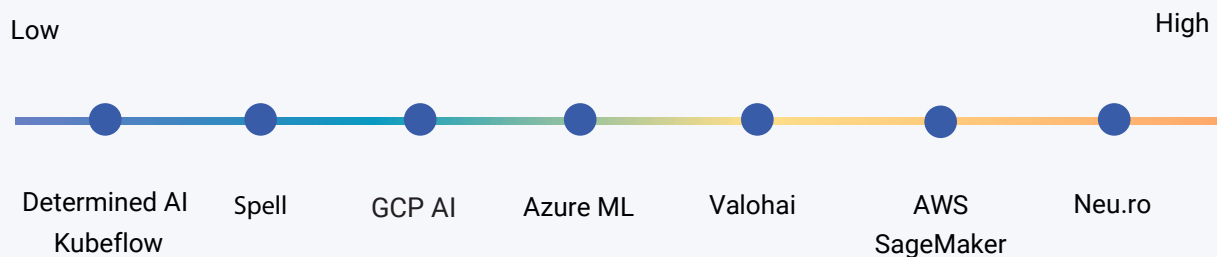
Gradient are at the tail of the list. Neither of them has any pre-made integrations. FloydHub has a documented CLI, but no API or SDK. Gradient supports both CLI and SDK, but no API. Both tools seem to assume a position of an all-in-one platform. They prioritize in-house feature development over third-party integrations and do not empower users to experiment with custom extensions.

Data Management Capabilities: Low to High

In 2020, **44%** of leaders in AI adoption used a standardized toolset to create production-ready data pipelines. Thus, we decided to analyze how different vendors deliver on that need. In particular, we assessed which native functionality (if any) the platform provides for building data ingestion pipelines and whether the vendor supports feature store and data registry creation.

- **Feature store** is a centralized storage unit for collecting, organizing, managing, and serving all feature values that were obtained from raw data.
- **Data registry** is a centralized storage unit for hosting different versions of datasets alongside meta information.

Data Management Capabilities: Low to High



Data Management Capabilities Rankings



Neu.ro turned out to be the leader of the pack. The platform supports integrations with different data registry backends. However, it doesn't support custom feature store development. If this option is critical for you, **AWS SageMaker** has more advanced feature store capabilities. The platform provides tools for building purpose-built repositories for storing and retrieving ML features. Other vendors on the list do not support feature store creation, but the Google AI platform has announced plans to deploy this functionality later in 2021. Data registry is another rarely supported feature, only provided as a managed service by Neu.ro.

In terms of data ingestion, **Azure Machine Learning** has the most diverse capabilities. With Azure, you have virtually no limits on data sources. However, you are prompted to use Azure Data Factory to set up data integration and transformations, plus configure Azure DevOps or GitHub for integration within web services and other services.

Valohai also received a high ranking as the platform lets you configure secure integrations with private cloud-based data repositories (AWS S3, Google Cloud Storage, Azure Blob Storage among others). A more expensive plan also includes support for on-premises data sources. **Determined AI** and **Kubeflow** offer no native capabilities for data management and require custom extensions. Respectively, they rank in the tail.

Model Development Experience: Low to High

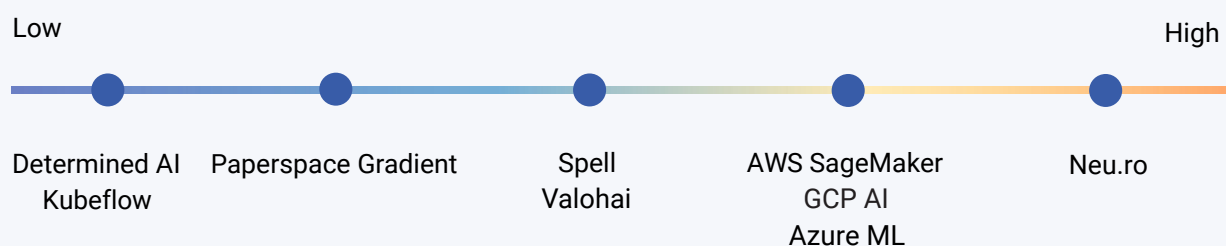
MLOps platforms differ drastically in terms of user-friendliness. Most were primarily designed for pro data scientists rather than business users or curious citizen developers. Respectively, most platforms come with a steep learning curve. Sometimes, that curve can be too challenging or time-consuming, even for seasoned coders who are forced to spend several hours on IDE setup and configuration every time they need to start a new experiment.

Hence, we decided to assess how vendors stack in terms of user experience. In particular, we looked at the following facets:

- Does the platform provide a convenient web UI? Any visual, drag-and-drop tools?
- What about APIs and CLIs to build custom integrations?
- Is there sufficient documentation available? Are interfaces documented as well (via references)?
- Do new users have any access to tutorials, reference architecture, pre-made recipes?
- How easy is it to start a new project and collaborate with other developers or business users?
- Does the platform support remote debugging?
- Is it easy to track and monitor different experiments?

Here's what we found

Model Development Experience: Low to High



Model Development Experience

Neu.ro and IaaS and proprietary platforms scored the highest in this regard. Open source projects lag behind in terms of development experience as many of them were never designed to support larger scaled collaborations.

The top-4 entrants score high in terms of supported interfaces — web UI, open APIs, and CLIs. Extensive documentation, tutorials, and recipes are easy to find as well. Most share GitHub repositories and support documentation publicly and/or provide a selection of Jupyter notebooks privately to users. **Gradient** scored lower than others in this category since the platform only supports CLI, but not APIs. The platform's web UI interfaces are very delightful though.

In terms of collaboration functionality, all proprietary vendors offer convenient and often customizable workspaces for productive teamwork. These provide a single-pane view into the current active experiments, model versions, and deployments, along with tools for monitoring, reproducing, and investigating different jobs.

When it comes to model development, **Neu.ro** and **Valohai** are the only two vendors offering project scaffolding — the ability to create a new project from a git-based project template. Neu.ro also comes with a user-friendly development environment (built with VSCode and Jupyter), a selection of pre-installed Jupyter Notebooks (based on your requests), a remote debugging toolkit, and experiment tracking capabilities powered by MLflow.

Valohai also lets you set up project templates and comes pre-packaged with a host of ready-to-use project recipes, libraries, and notebooks. The core difference between the two is that Valohai does not allow you to connect a custom debugger, whereas you can run one with Neu.ro.

AWS, Google, and Azure also provide a familiar and convenient IDE for model development but do not feature any project templates or recipes. **Spell, Gradient** encourage you to experiment by including a small number of easy-to-replicate sample projects. But none of them support remote debugging — that’s a constraint all proprietary MLOps platforms share.

Determined AI, Gradient, and Spell also somewhat lag when it comes to experimenting tracking. While every vendor lets you track model performance results and changes over time, they don’t provide an opportunity to benchmark different models or compare two different experiments at a great level of detail.

Model Training and Management Capabilities: Low to High

Almost a half (48%) of AI leaders rely on automated tools for developing and testing new models. Automated model training pipelines, hyperparameter tuning, and distributed training can accelerate experimentation speed and quality. Model management functionality, in turn, helps ensure that successfully trained models won't crash in production due to misconfigurations or other Ops mishaps.

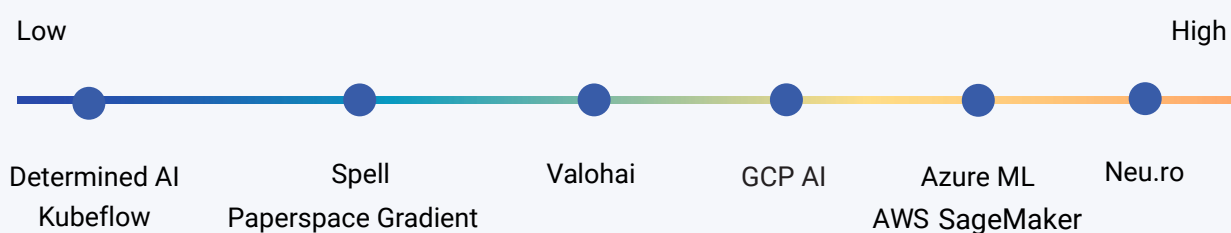
For that reason, we specifically looked at how different MLOps vendors facilitate model training and whether they provide pre-made training pipelines, hyperparameter tuning, and model registry creation.

Hyperparameter tuning (optimization) stands for different approaches to selecting the optimal model parameters, whose values will be used to control the model learning process.

Model registry is a storage service for collecting, managing, and tracking different model artifacts (and other metadata), required for successful model deployments.

Automation pipelines transpose CI/CD principles to model development and automate jobs scheduling, execution, and orchestration. Pipelines introduce a greater degree of standardization into the machine learning lifecycle and enable faster interactions and more predictable deployments.

Model Training and Management Capabilities: Low vs High



Model Training and Management Rankings

Neu.ro scored the highest among other comparable platforms since it ticks all four boxes in terms of training and model management features:

- Distributed training
- Hyperparameter tuning
- Automation (pipelines)
- Model registry

The platform provides access to an array of ready-to-use feature sets for model training, re-training, and validation. Furthermore, you can request any custom configurations you need. They provide a Neural Network Intelligence (NNI) toolkit, developed by Microsoft and distributed as open-source for hyperparameter optimization.

Azure ML comes in a close second due to somewhat more limited model registry functionality available as part of the Azure Machine Learning Studio. **AWS SageMaker** recently added training and deployment pipelines to their mix of features and also supports model registry setup. You can conveniently catalog models for production, manage versions, record relevant metadata, and then automate model deployment with CI/CD. On the other hand, **Google AI platform** scores exceptionally well in all areas except for hyperparameter tuning, though they are constantly expanding the list of supported parameters.

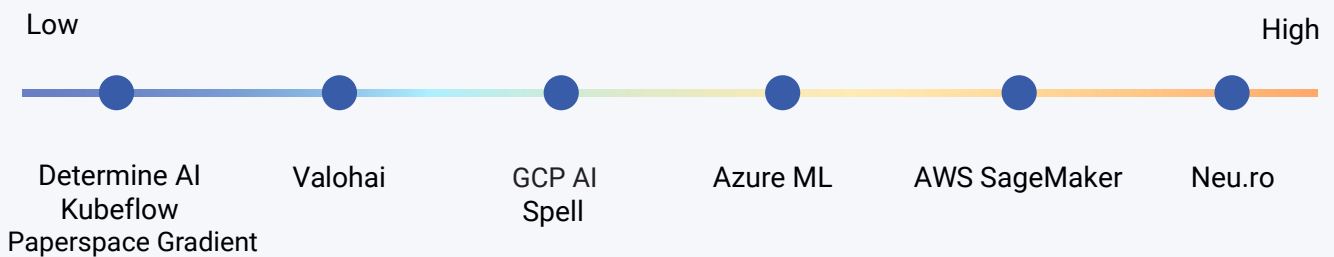
Spell lag in terms of training automation pipelines. **Gradient** has pipelines in preview, but we didn't yet have the chance to test the announced functionality. Neither of them supports model registry setup. **Valohai**, who doesn't have this option either, but compensates for the lack of it with robust training pipelines. The platform has mature pipeline management capabilities and supports a decent variety of automation scenarios — from data pre-processing to hyperparameter sweeps and model deployments.

Model Deployment and Monitoring: Low to High Difficulty

Back in 2018, over **70%** of data scientists' productive time was spent on model deployments. In 2020, the figure dropped to 25%, most likely due to the arrival of (semi-) automated deployment pipelines for ML and the broader adoption of containerized model deployments. On the other hand, the shift may have also happened because data scientists handed off model deployments to Ops teams.

Still, "hand-offs" without proper tools are not a good recipe for frequent, scalable, and stable model deployments. But MLOps adoption is improving since many vendors now provide pre-made model deployment pipelines built based on the CI/CD principles. Additionally, some market leaders also provide value-added functionality such as model monitoring and model explainability tools for tracking model performance over time.

Model Deployment and Monitoring: Low vs High Difficulty



Model Deployment and Monitoring Rankings

Neu.ro MLOps platform scored the highest in this category since it integrates with popular open-source tools for model deployments (Seldon Core) and monitoring (Prometheus + Grafana). **AWS SageMaker** also provides a native Model Monitor tool, but model drift observations are limited to data quality, accuracy, bias, and feature attribution. Whereas you can track and visualize more parameters with Prometheus + Grafana combo.

Azure Machine Learning has limited monitoring capabilities. When it comes to model performance monitoring, you can only:

- Observe the model's serving data for drift using the model's training data as a proxy for accuracy
- Track time-series datasets for drift from a previous period of time

Google AI platform doesn't provide any tools for monitoring deployments but has moderately good tools for registering drifts in performance. At the beginning of 2021, Spell announced a partnership with Arize. The platform's customers can set up an integration to monitor model performance in production and receive timely insights.

Determine AI, **Kubeflow**, and **Gradient** do not offer any native capabilities for deployments or model observability.

Startup vs Enterprise-Grade MLOps Platforms

Thanks to open-source solutions, machine learning, and deep learning have become more or less commoditized technologies. College students and senior AI engineers all stick to the same ML lifecycle when designing new models and running experiments.

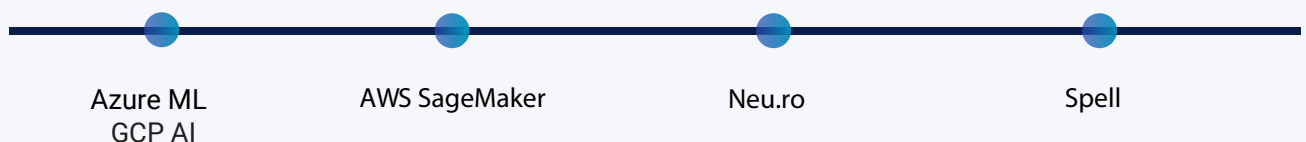
However, different groups of users run projects of varying scale and complexity. They also have varying requirements when it comes to model audits, explainability, and resources management.

To determine how different vendors compare in terms of scaling capabilities, we analyzed the scope of supported features commonly required by enterprise users. These include:

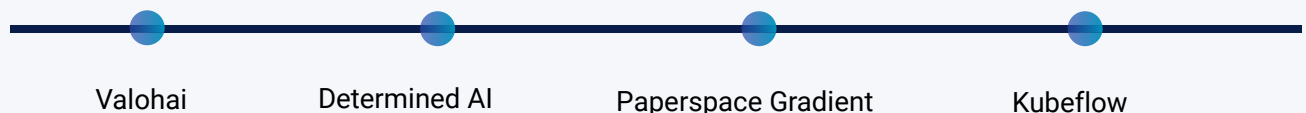
- Access control and identity management for different projects
- Secure Single Sign-On (SSO) to minimize the frictions in credentials management
- Log audit capabilities for security
- Resources monitoring and consumption reports

Startup vs Enterprise-Grade MLOps Platforms

Best for enterprise projects



Best for smaller projects



Unsurprisingly, cloud vendors rank best in this category. That makes sense given their vast experience with servicing enterprise clients and general orientation towards this customer segment. The only area where cloud vendors lag is SSO — most do not support social credentials or logins via other platforms except for GitHub. **AWS SageMaker** also received a lower ranking for the reason since it doesn't support GitHub logins.

Otherwise, the trio is well-stacked in resources monitoring tools, audit capabilities, access control management, and overall infrastructure scalability to accommodate more extensive ML operations and multiple deployments.

Neu.ro share the third position. Both support SSO and provide detailed resource usage analytics. However, Floydhub only does so for their proprietary hardware, whereas Neu.ro gives analytics for all connected GPUs/CPUs. The two platforms have more limited access control functionality than the leaders and do not provide log audits.

Spell was ranked lower on the list due to the lack of audit capabilities and reports on resource consumption. They do, however, provide a stellar toolkit for provisioning new instances in seconds across multiple projects and monitoring resource usage. SSO is only limited to the most expensive custom enterprise plans. Other entrants of the list have none of those mentioned above functionality, which can be a deal-breaker for many enterprise clients.

Conclusions

In this report, we attempted to coax clarity out of chaos when it comes to MLOps platform functionality, positioning, and ability to deliver on the advertised functionality. Overall, the report features a diverse range of established players: IaaS providers, rapidly evolving challengers, and niche players. We also only evaluated platforms with commercially licensable products and commercially licensed open-sources platforms. It's important to note, however, that most platforms on the list rely on or integrate with open-sources tools, frameworks, and libraries, making them even more robust.

Ultimately, when it comes to your choice of vendor, be sure to conduct an independent assessment and assess each MLOps platform through the lens of corporate requirements. Engage your data science team in the vendor selection process. Collect their feedback and commentary on ongoing struggles and prioritize them in a list of requirements. Then seek out a vendor that ticks the most of these boxes.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material
for any purpose, even commercially.

Attribution — You must give appropriate credit and link to the creator of this report neu.ro

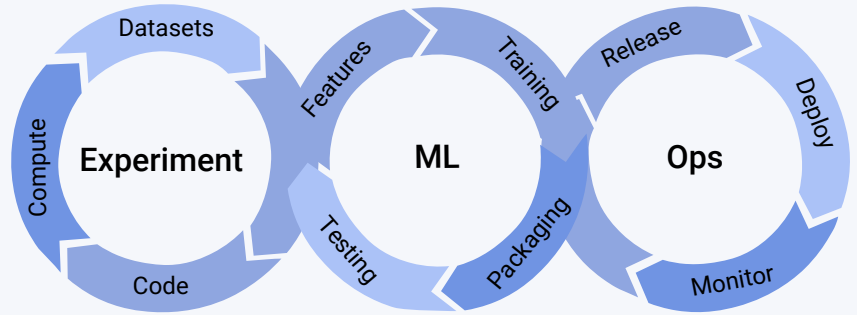


The Growing MLOps Market

Main Vendors

IDEs revolutionized the speed and quality of software development. MLOps platforms attempt to create a similar level of speed, convenience, and automation for data scientists with the ultimate promise of reducing model development and deployment timelines by 2X-10X.

Machine Learning Lifecycle



50%



Of AI experts use standard development tools and frameworks to create AI models

[McKinsey 2020](#)

73%



Of business leaders believe that MLOps adoption will keep them competitive

[Forrester 2020](#)

\$4 bln

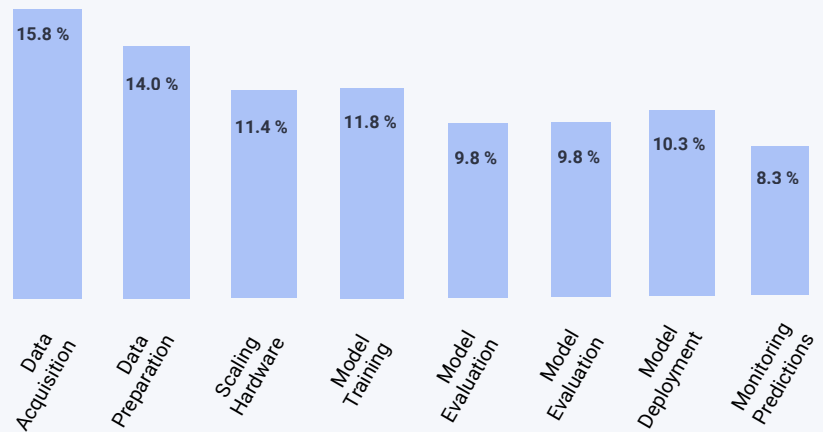
In annual revenue is projected to be generated by MLOps platforms by 2025

[Deloitte 2021](#)

1 to 30 days

42% of AI leaders require to deploy a new model

[Algorithmia 2020](#)

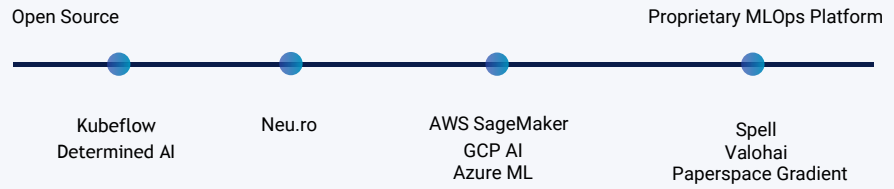


MLOps Market Growth Prospects

The MLOps market was estimated at \$23.2 billion in 2019 and is projected to reach \$126 billion by 2025 due to rapid adoption



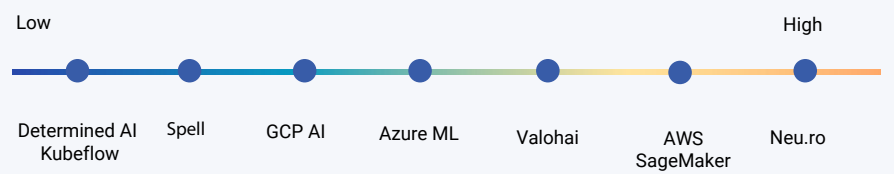
Open Source vs Proprietary MLOps Platform



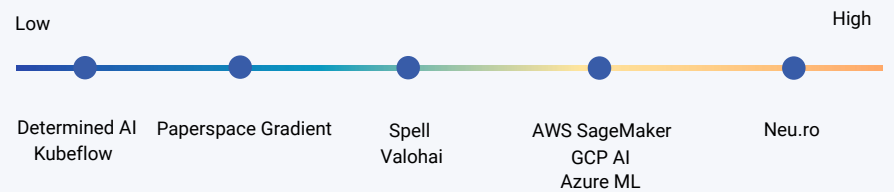
Interoperability: Low to High



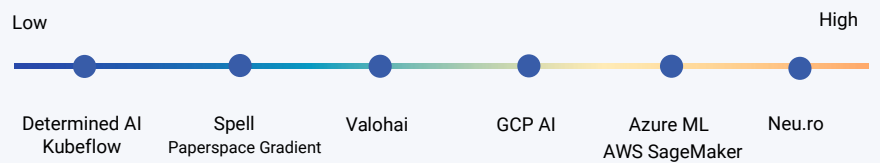
Data Management Capabilities: Low to High



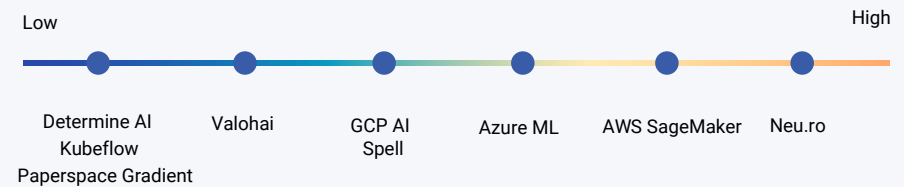
Model Development Experience: Low to High



Model Training and Management Capabilities: Low to High

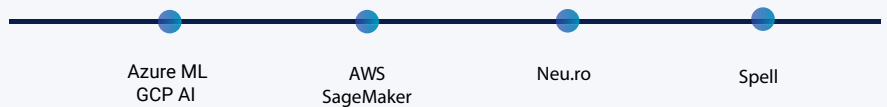


Model Deployment and Monitoring: Low vs High Difficulty



Enterprise-Grade vs Startup-Grade MLOps Platforms

Best for enterprise



Best for small projects

