Machine Learning and Large Language Model Operations Buyers Guide

Software Provider and Product Assessment





Machine Learning and Large Language Model Operations

Artificial intelligence (AI) has continued to rise in prominence. ISG Market Lens Research shows that 79% of enterprises plan to increase their spending on AI technology and that the growth in AI spending is outpacing the growth in all other categories of IT spending. However, the research also shows that only a fraction of AI applications are in production. The process of developing and deploying AI applications involves multiple interrelated and complicated steps. In addition, enterprises are grappling with ways to ensure AI applications comply with internal governance policies and an evolving external regulatory environment. One of the ways enterprises can improve the process of moving AI applications to production is through machine learning and large language model operations (ML/LLMOps).

ISG Research defines ML/LLMOps as the processes used to develop, deploy, monitor, manage and govern ML and LLMs. Developing and deploying AI models is a multistep process, beginning with collecting and curating the data that will be used to create the model. Once a model is developed and tuned using the training data, it needs to be tested to determine its

ISG Research defines ML/LLMOps as the processes used to develop, deploy, monitor, manage and govern ML and LLMs. accuracy and performance. Then the model needs to be applied in an operational application or process. For example, in a customer service application, a predictive AI model might make a recommendation for how a representative should respond to the customer's situation. Similarly, a self-service customer application might use an LLM to provide a chatbot or guided experience to deliver those recommendations.

For many years, data scientists, data engineers and development operations (DevOps) teams needed to cobble together processes to support Al/ML models. Each team operated largely independently of the other,

and there was no sense that the various steps were connected to one another. As AI grew in popularity and as models were updated more frequently to reflect changing market conditions, these ad hoc processes became an obstacle to scaling and governing AI. The notions of repeatability and automation had become common themes in DevOps. The AI community began to apply some of these same concepts to AI processes, eventually referring to them as machine learning operations or MLOps.

Data is flowing throughout these processes. Considerable time and effort are invested in preparing data to feed into predictive models. Feature engineering requires exploration and experimentation with the data. Once the features are identified, robust repeatable processes are needed to create data pipelines that feed these features into the models. In the case of generative AI (GenAI), data—often in the form of documents—feeds custom LLM development or fine-tuning. Additional data flows through the prompting process to direct LLMs to provide

more specific and more accurate responses. Enterprises must govern these data flows to ensure compliance with internal policies and regulatory requirements. The regulatory environment is emerging and evolving, with the European Union passing the AI Act, the U.S. issuing and then rescinding an Executive Order on responsible development of AI, and dozens of U.S. states either enacting or proposing AI regulations.

The process does not conclude once a model is deployed. Enterprises need to monitor and maintain the models, ensuring they continue to be accurate and relevant as market conditions change. Realistically, it is only a matter of time before a model's accuracy declines to the point where it can be replaced by another, more accurate model. The new model may simply be the result of retraining the old model on new data or it may be the result of using different modeling techniques. In either case, the models must be monitored constantly and updated as and when necessary. In the case of third-party LLMs, the providers are constantly updating and improving their models, so enterprises need to be prepared to deploy the newer models as well.

LLMs introduce additional operational concerns. Both prompts and responses must be monitored. In the case of prompts, enterprises need to ensure sensitive information or customers' private data is not being shared with third parties. Prompts must also be monitored to prevent prompt injection to avoid hijacking. Responses must be monitored for toxicity and bias. Enterprises also want to be able to trace or explain the responses being generated to increase trust and understanding in the use of Al.

Software providers slowly recognized that the lack of ML/LLMOps tooling was inhibiting successful use of AI. Enterprises were left to their own devices to create scripts and piece

together solutions to address these issues. Fortunately, AI software providers have expanded their platforms to address many of these capabilities, and specialist providers have emerged with a focus on MLOps/LLMOps. In fact, we assert that by 2027, 4 in 5 enterprises will use MLOps and LLMOps tools to improve the quality and governance of their AI/ML efforts.

All of these capabilities are important to maximize the success of AI investments. As a result, our evaluation of AI software providers considers each of them. Our separate AI Platforms Buyers Guide includes a superset of AI functionality, including ML/LLMOps. Our Generative and



Agentic Al Buyers Guide examines the subset of functionality to support the development and use of LLMs and agent processes. This Buyers Guide focuses on the ML/LLMOps functionality described above. While most software providers offer a combination of capabilities, many have evolved from one specific segment of the market or another. As a result, they tend to be

more capable in the segment from which they originated. This Buyers Guide helps enterprises identify the relative strengths of providers in each segment of the market.

The ML/LLMOps segment of the market has been evolving rapidly to meet the needs of enterprises. Products may be incomplete in one area or another. Our previous Buyers Guide indicated that less than one-quarter of software providers fully met enterprise requirements for data governance or repeatability. Only 1 in 5 provided automatic documentation of models or had adequate approval processes to support new model deployment. As a result, enterprises should expect to supplement software-based ML/LLMOps with other processes to ensure they are meeting their internal and external compliance requirements. ISG Market Lens Research indicates that the most common thing enterprises would do differently is better coordination and governance of their Al implementations.

The ISG Buyers Guide[™] for Machine Learning and Large Language Model Operations only evaluates software providers and products with specific ML/LLMOps support. The ML/LLMOps Buyers Guide uses portions the AI platform capability framework, and to be included in this Buyers Guide, products must include AI/ML pipelines, LLM fine-tuning processes, developer tooling, repeatability, monitoring, governance and deployment capabilities. The capability model evaluated includes: advanced model optimization, data preparation, developer and data scientist tooling, generative AI, MLOps, and types of AI/ML modeling. All of these capabilities are critical to ensure that enterprises can operationalize and rely upon the models that are produced in their AI processes.

This research evaluates the following software providers that offer products that address key elements of MLOps and LLMOps as we define it: Alibaba Cloud, Altair, Alteryx, Anaconda, AWS, C3 Al, Cloudera, Databricks, Dataiku, DataRobot, Domino Data Lab, Google Cloud, H2O.ai, Huawei Cloud, Hugging Face, IBM, MathWorks, Microsoft, NVIDIA, Oracle, Palantir, Quantexa, Red Hat, Salesforce, SAP, SAS, Snowflake, Teradata and Weights & Biases.

Buyers Guide Overview

For over two decades, ISG Research has conducted market research in a spectrum of areas across business applications, tools and technologies. We have designed the Buyers Guide to provide a balanced perspective of software providers and products that is rooted in an understanding of the business requirements in any enterprise. Utilization of our research

ISG Research has designed the Buyers Guide to provide a balanced perspective of software providers and products that is rooted in an understanding of business requirements in any enterprise. methodology and decades of experience enables our Buyers Guide to be an effective method to assess and select software providers and products. The findings of this research undertaking contribute to our comprehensive approach to rating software providers in a manner that is based on the assessments completed by an enterprise.

The ISG Buyers Guide[™] for Machine Learning and Large Language Model Operations is the distillation of over a year of market and product research efforts. It is an assessment of how well software providers' offerings address enterprises' requirements for ML and LLMOps software. The index is structured to support a request for information (RFI) that could be used in the request for proposal (RFP) process by incorporating all criteria needed to evaluate, select, utilize and maintain relationships with software providers. An effective product and customer experience with a provider can ensure the best long-term relationship and value achieved from a resource and financial investment.

In this Buyers Guide, ISG Research evaluates the software in seven key categories that are weighted to reflect buyers' needs based on our expertise and research. Five are product-experience related: Adaptability, Capability, Manageability, Reliability, and Usability. In addition, we consider two customer-experience categories: Validation, and Total Cost of Ownership/Return on Investment (TCO/ROI). To assess functionality, one of the components of Capability, we applied the ISG Research Value Index methodology and blueprint, which links the personas and processes for ML and LLMOps to an enterprise's requirements.

The structure of the research reflects our understanding that the effective evaluation of software providers and products involves far more than just examining product features, potential revenue or customers generated from a provider's marketing and sales efforts. We believe it is important to take a comprehensive, research-based approach, since making the wrong choice of ML and LLMOps technology can raise the total cost of ownership, lower the return on investment and hamper an enterprise's ability to reach its full performance potential. In addition, this approach can reduce the project's development and deployment

time and eliminate the risk of relying on a short list of software providers that does not represent a best fit for your enterprise.

ISG Research believes that an objective review of software providers and products is a critical business strategy for the adoption and implementation of ML and LLMOps software and applications. An enterprise's review should include a thorough analysis of both what is possible and what is relevant. We urge enterprises to do a thorough job of evaluating ML and LLMOps systems and tools and offer this Buyers Guide as both the results of our in-depth analysis of these providers and as an evaluation methodology.

How To Use This Buyers Guide

Evaluating Software Providers: The Process

We recommend using the Buyers Guide to assess and evaluate new or existing software providers for your enterprise. The market research can be used as an evaluation framework to establish a formal request for information from providers on products and customer experience and will shorten the cycle time when creating an RFI. The steps listed below provide a process that can facilitate best possible outcomes.

1. <u>Define the business case and goals.</u>

Define the mission and business case for investment and the expected outcomes from your organizational and technological efforts.

- Specify the business needs.
 Defining the business requirements helps identify what specific capabilities are required with respect to people, processes, information and technology.
- Assess the required roles and responsibilities.
 Identify the individuals required for success at every level of the enterprise from executives to frontline workers and determine the needs of each.
- <u>Outline the project's critical path.</u>
 What needs to be done, in what order and who will do it? This outline should make clear the prior dependencies at each step of the project plan.
- <u>Ascertain the technology approach.</u>
 Determine the business and technology approach that most closely aligns to your enterprise's requirements.
- <u>Establish software provider evaluation criteria.</u>
 Utilize the product experience: Adaptability, Capability, Manageability, Reliability and Usability, and the customer experience in TCO/ROI and Validation.
- Evaluate and select the technology properly.
 Weight the categories in the technology evaluation criteria to reflect your enterprise's priorities to determine the short list of software providers and products.
- Establish the business initiative team to start the project.
 Identify who will lead the project and the members of the team needed to plan and execute it with timelines, priorities and resources.

The Findings

All of the products we evaluated are feature-rich, but not all the capabilities offered by a software provider are equally valuable to types of workers or support everything needed to manage products on a continuous basis. Moreover, the existence of too many capabilities

may be a negative factor for an enterprise if it introduces unnecessary complexity. Nonetheless, you may decide that a larger number of features in the product is a plus, especially if some of them match your enterprise's established practices or support an initiative that is driving the purchase of new software.

Factors beyond features and functions or software provider assessments may become a deciding factor. For example, an enterprise may face budget constraints such that the TCO evaluation can tip the balance to one provider or another. This is where the Value Index methodology and the appropriate category weighting can be applied to determine the best fit of software providers and products to your specific needs.

Overall Scoring of Software Providers Across Categories

The research finds Oracle atop the list, followed by AWS and Databricks. Providers that place in the top three of a category earn the designation of Leader. Oracle has done so in seven categories, Databricks in four, Google Cloud and Microsoft in three, AWS in two and Dataiku and Teradata in one category.

Providers	Grade	Performance	
Oracle	A-	Leader 82.1%	
AWS	B++	Leader 79.9%	
Databricks	B++	Leader 79.0%	
Microsoft	B++	77.1%	
Google Cloud	B++	76.4%	
IBM	B++	76.2%	
Dataiku	B++	75.1%	
DataRobot	B+	72.2%	
Teradata	B+	71.3%	
Domino Data Lab	B+	70.6%	
Alteryx	В	68.4%	
Cloudera	В	67.3%	
Snowflake	В	67.3%	
MathWorks	В	65.2%	
Salesforce	В	64.9%	
Alibaba Cloud	В	64.5%	
H2O.ai	В	64.0%	
SAS	В	63.5%	
Palantir	В	63.3%	
SAP	В	63.3%	
VIDIA	B-	60.9%	
Huawei Cloud	B-	60.4%	
Altair	B-	60.3%	
Hugging Face	B-	59.0%	
Weights & Biases	B-	59.0%	
Red Hat	B-	57.6%	
Anaconda	B-	56.6%	
C3 AI	C++	53.5%	
Quantexa	C+	45.7%	

The overall representation of the research below places the rating of the Product Experience and Customer Experience on the *x* and *y* axes, respectively, to provide a visual representation and classification of the software providers. Those providers whose Product Experience have a higher weighted performance to the axis in aggregate of the five product categories place farther to the right, while the performance and weighting for the two Customer Experience categories determines placement on the vertical axis. In short, software providers that place closer to the upper-right on this chart performed better than those closer to the lower-left. The research places software providers into one of four overall categories: Assurance, Exemplary, Merit or Innovative. This representation classifies providers' overall weighted performance.



Exemplary: The categorization and placement of software providers in Exemplary (upper right) represent those that performed the best in meeting the overall Product and Customer Experience requirements. The providers rated Exemplary are: Alteryx, AWS, Cloudera, Databricks, Dataiku, Domino Data Lab, Google Cloud, IBM, Microsoft, Oracle, SAS, Snowflake and Teradata.

Innovative: The categorization and placement of software providers in Innovative (lower right) represent those that performed the best in meeting the overall Product Experience requirements but did not achieve the highest levels of requirements in Customer Experience. The providers rated Innovative are: Alibaba Cloud, DataRobot, H2O.ai and MathWorks.

Assurance: The categorization and placement of software providers in Assurance (upper left) represent those that achieved the highest levels in the overall Customer Experience requirements but did not achieve the highest levels of Product Experience. The providers rated Assurance are: NVIDIA, Red Hat, Salesforce and SAP. **Merit**: The categorization of software providers in Merit (lower left) represents those that did not exceed the median of performance in Customer or Product Experience or surpass the threshold for the other three categories. The providers rated Merit are: Altair, Anaconda, C3 AI, Huawei Cloud, Hugging Face, Palantir, Quantexa and Weights & Biases.

We warn that close provider placement proximity should not be taken to imply that the packages evaluated are functionally identical or equally well suited for use by every enterprise or for a specific process. Although there is a high degree of commonality in how enterprises handle ML and LLMOps, there are many idiosyncrasies and differences in how they do these functions that can make one software provider's offering a better fit than another's for a particular enterprise's needs.

We advise enterprises to assess and evaluate software providers based on organizational requirements and use this research as a supplement to internal evaluation of a provider and products.

Product Experience

The process of researching products to address an enterprise's needs should be comprehensive. Our Value Index methodology examines Product Experience and how it aligns with an enterprise's life cycle of onboarding, configuration, operations, usage and maintenance. Too often, software providers are not evaluated for the entirety of the product; instead, they are evaluated on market execution and vision of the future, which are flawed since they do not represent an enterprise's requirements but how the provider operates. As more software providers orient to a complete product experience, evaluations will be more robust.

The research results in Product Experience are ranked at 80%, or four-fifths, of the overall rating using the specific underlying weighted category performance. Importance was placed on the categories as follows: Usability (10%), Capability (40%), Reliability (10%), Adaptability (10%) and Manageability (10%). This weighting impacted the resulting overall ratings in this research. Oracle, AWS and Microsoft were designated Product Experience Leaders. While not Leaders, Dataiku and Databricks were also found to meet a broad range of enterprise product experience requirements.

ML & LLM Operations Product Experience

Providers	Grade	Performance
Oracle	B++	Leader 63.0%
AWS	B++	Leader 62.7%
Microsoft	B+	Leader 59.7%
Dataiku	B+	59.6%
Databricks	B+	59.6%
BM	B+	58.3%
Google Cloud	B+	57.9%
DataRobot	B+	56.6%
Domino Data Lab	В	53.4%
Teradata	В	52.5%
MathWorks	В	50.7%
H2O.ai	В	50.4%
Cloudera	В	50.3%
lteryx	B-	50.0%
nowflake	B-	49.8%
libaba Cloud	B-	47.7%
AS	B-	47.4%
lugging Face	B-	46.6%
alantir	B-	46.6%
alesforce	B-	46.1%
Veights & Biases	B-	45.6%
ltair	B-	45.1%
AP	C++	44.3%
luawei Cloud	C++	43.9%
IVIDIA	C++	43.3%
naconda	C++	42.2%
ed Hat	C++	40.6%
3 AI	C+	38.6%
Quantexa	С	31.0%

Customer Experience

The importance of a customer relationship with a software provider is essential to the actual success of the products and technology. The advancement of the Customer Experience and the entire life cycle an enterprise has with its software provider is critical for ensuring

satisfaction in working with that provider. Technology providers that have chief customer officers are more likely to have greater investments in the customer relationship and focus more on their success. These leaders also need to take responsibility for ensuring this commitment is made abundantly clear on the website and in the buying process and customer journey.

The research results in Customer Experience are ranked at 20%, or one-fifth, using the specific underlying weighted category performance as it relates to the framework of commitment and value to the software provider-customer relationship. The two evaluation categories are Validation (10%) and TCO/ROI (10%), which are weighted to represent their importance to the overall research.

The software providers that evaluated the highest overall in the aggregated and weighted Customer Experience categories are Databricks, Oracle and Google Cloud. These category leaders best communicate commitment and dedication to customer needs. While not Leaders, IBM and Teradata were also found to meet a broad range of enterprise customer experience requirements.

ML & LLM Operations Customer Experience

Providers	Grade	Performance
Databricks	Α	Leader 17.7%
Oracle	Α	Leader 17.6%
Google Cloud	A-	Leader 16.6%
IBM	A-	16.5%
Teradata	A-	16.4%
AWS	A-	16.3%
Alteryx	B++	16.0%
Domino Data Lab	B++	15.9%
Microsoft	B++	15.8%
Salesforce	B++	15.7%
NVIDIA	B++	15.5%
SAP	B++	15.4%
Dataiku	B++	15.2%
Snowflake	B++	15.0%
Red Hat	B+	14.9%
Cloudera	B+	14.8%
SAS	B+	14.7%
DataRobot	B+	14.6%
Alibaba Cloud	B+	14.5%
Palantir	B+	14.4%
Huawei Cloud	B+	13.9%
C3 AI	B+	13.8%
Altair	В	13.7%
MathWorks	В	13.4%
H2O.ai	В	13.1%
Anaconda	В	13.0%
Weights & Biases	В	12.6%
Quantexa	B-	12.5%
Hugging Face	B-	11.6%
Hugging Face Source: ISG Resea ML & LLM Operat © 2025 Informatic	B- rch ions Buyer	s Guide Group, Inc.

Software providers that did not perform well in this category were unable to provide sufficient customer case studies to demonstrate success or articulate their commitment to customer experience and an enterprise's journey. The selection of a software provider means a continuous investment by the enterprise, so a holistic evaluation must include examination of how they support their customer experience.

Appendix: Software Provider Inclusion

For inclusion in the ISG Buyers Guide[™] for Machine Learning and Large Language Operations in 2025, a software provider must be in good standing financially and ethically, have at least \$25 million in annual or projected revenue verified using independent sources, sell products and provide support on at least two continents and have at least 50 customers. The principal source of the relevant business unit's revenue must be software-related, and there must have been at least one major software release in the past 12 months.

To be included in the Machine Learning and Large Language Operations Buyers Guide, the software provider must enable AI/ML pipelines, LLM fine-tuning processes, developer tooling, repeatability, monitoring, governance and deployment capabilities. The product(s) must be actively marketed as providing machine learning and large language operations functionality.

The research is designed to be independent of the specifics of software provider packaging and pricing. To represent the real-world environment in which businesses operate, we include providers that offer suites or packages of products that may include relevant individual modules or applications. If a software provider is actively marketing, selling and developing a product for the general market and it is reflected on the provider's website that the product is within the scope of the research, that provider is automatically evaluated for inclusion.

All software providers that offer relevant ML and LLMOps products and meet the inclusion requirements were invited to participate in the evaluation process at no cost to them.

Software providers that meet our inclusion criteria but did not completely participate in our Buyers Guide were assessed solely on publicly available information. As this could have a significant impact on classification and ratings, we recommend additional scrutiny when evaluating those providers.

Products Evaluated

Provider	Provider Product Names		Release Month/Year	
	Platform for Al	2025-03-28	March 2025	
Alibaba Cloud	Model Studio	N/A	March 2025	
Altair	Al Hub	2025.0	February 2025	
Altali	Al Studio	2025.0	December 2024	
Alteryx	Analytics Cloud Platform	N/A	December 2024	
Anaconda	Al Platform	5.8.1.3	March 2025	
AWS	Amazon SageMaker Al	N/A	March 2025	
C2 A1	C3 Agentic Al Platform	N/A	April 2025	
C3 AI	C3 Generative Al	N/A	April 2025	
Cloudera	Cloudera Al	2.0.50-b50 (on- prem 1.5.4)	March 2025	
Databricks	Mosaic Al	N/A	April 2025	
Dataiku	Dataiku	13.5.0	April 2025	
DataRobot	Enterprise Al Suite	11.0.1	April 2025	
Domino Data Lab	Domino Domino Cloud	6.0.4 N/A	April 2025 April 2025	
Google Cloud	Vertex Al Platform	1	April 2025	
H2O.ai	H2O AI Cloud	25.04.0	April 2025	
Huawei Cloud	ModelArts	6.8.0-HCS	March 2025	
Hugging Face	Enterprise Hub	0.31.0	April 2025	
IBM	watsonx.ai	2.1.2	March 2025	
MathWorks	MATLAB	R2024b	April 2025	
Microsoft	Azure Al Foundry	N/A	April 2025	
NVIDIA	NVIDIA AI Workbench	2025.04.01	April 2025	

Oracle	Oracle Cloud Infrastructure Generative Al Oracle Cloud Infrastructure Generative Al Agents Oracle Cloud Infrastructure Data Science	N/A N/A N/A	April 2025 April 2025 April 2025	
Palantir	AIP	N/A	April 2025	
Quantexa	Decision Intelligence Platform	2.7	April 2025	
Red Hat	Red Hat OpenShift Al	N/A	April 2025	
Salesforce	Salesforce Data Cloud Agentforce		February 2025 February 2025	
SAP	SAP SAP Business AI (Joule) SAP AI Core		February 2025 April 2025	
SAS	SAS SAS Viya		April 2025	
Snowflake	Snowflake The Snowflake Platform		April 2025	
Teradata	Teradata Teradata VantageCloud		April 2025	
Weights & Biases	W&B Weave W&B Models	0.68 0.68	April 2025 April 2025	

Providers of Promise

We did not include software providers that, as a result of our research and analysis, did not satisfy the criteria for inclusion in this Buyers Guide. These are listed below as "Providers of Promise."

Provider	Product	Revenue	Capability	International	Customers
Clarifai	Clarifai	No	Yes	Yes	Yes
Crowdworks	Crowdworks Al	Yes	Yes	No	Yes
DeepSeek	DeepSeek Platform	No	Yes	No	Yes
EdgeVerve	Al Next	Yes	No	Yes	Yes
FICO	FICO Platform	Yes	No	Yes	Yes
KNIME	KNIME Analytics Platform	No	Yes	Yes	Yes
Pecan	Pecan	No	Yes	Yes	Yes

About ISG Software Research and Advisory

ISG Software Research and Advisory provides market research and coverage of the technology industry, informing enterprises, software and service providers, and investment firms. The ISG Buyers Guides provide insight on software categories and providers that can be used in the RFI/RFP process to assess, evaluate and select software providers.

About ISG Research

ISG Research provides subscription research, advisory, consulting and executive event services focused on market trends and disruptive technologies. ISG Research delivers guidance that helps businesses accelerate growth and create more value. For further information about ISG Research subscriptions, please visit <u>research.isg-one.com</u>.

About ISG

ISG (Nasdaq: III) is a global Al-centered technology research and advisory firm. A trusted partner to more than 900 clients, including 75 of the world's top 100 enterprises, ISG is a long-time leader in technology and business services sourcing that is now at the forefront of leveraging Al to help organizations achieve operational excellence and faster growth. The firm, founded in 2006, is known for its proprietary market data, in-depth knowledge of provider ecosystems, and the expertise of its 1,600 professionals worldwide working together to help clients maximize the value of their technology investments.