White Paper

Five Questions Every Business Should Ask When Choosing an Enterprise-Grade Machine Learning Platform

Sponsored by: Databricks

Ritu Jyoti
August 2022

INTRODUCTION

Artificial intelligence/machine learning (AI/ML) remains a foundational investment in digital transformation initiatives. Organizations across the globe are adopting AI/ML technologies to improve operational efficiency, drive innovation, and improve employer productivity. Primary use cases for AI/ML technologies include IT automation, fraud/threat detection, and intelligent process automation (see Figure 1).

FIGURE 1

AI/ML Use Cases

Q.   What are the main use cases for which your organization is currently investigating, building, or deploying AI applications/solutions?
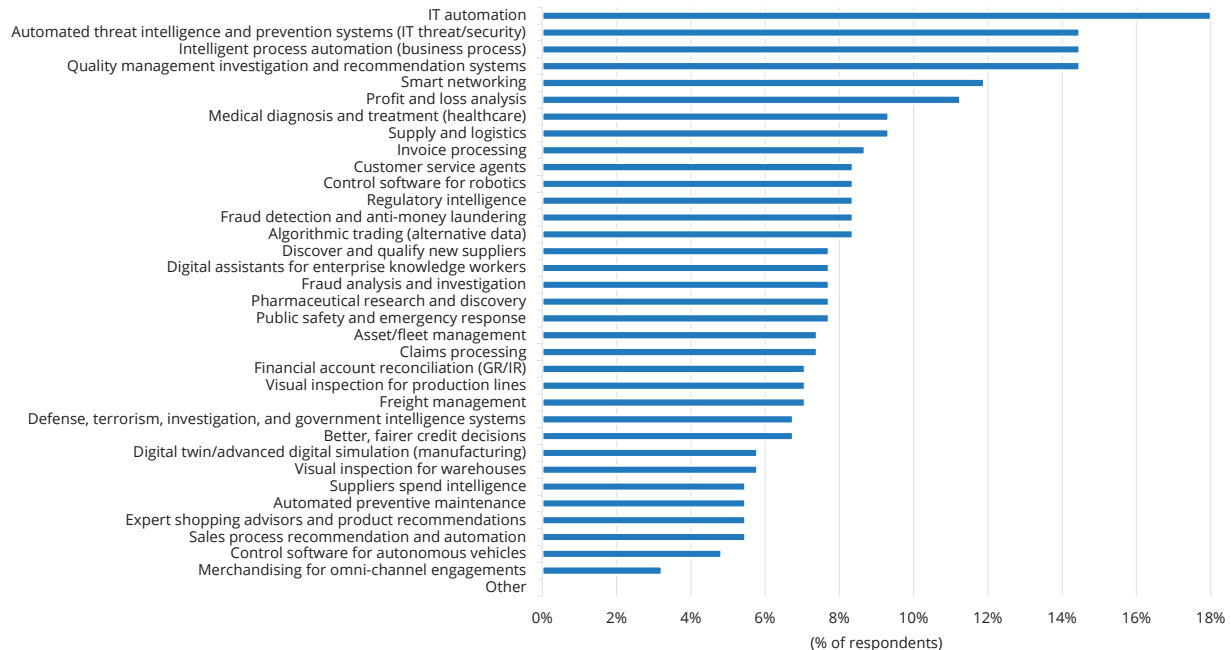
n = 311

Base = all respondents

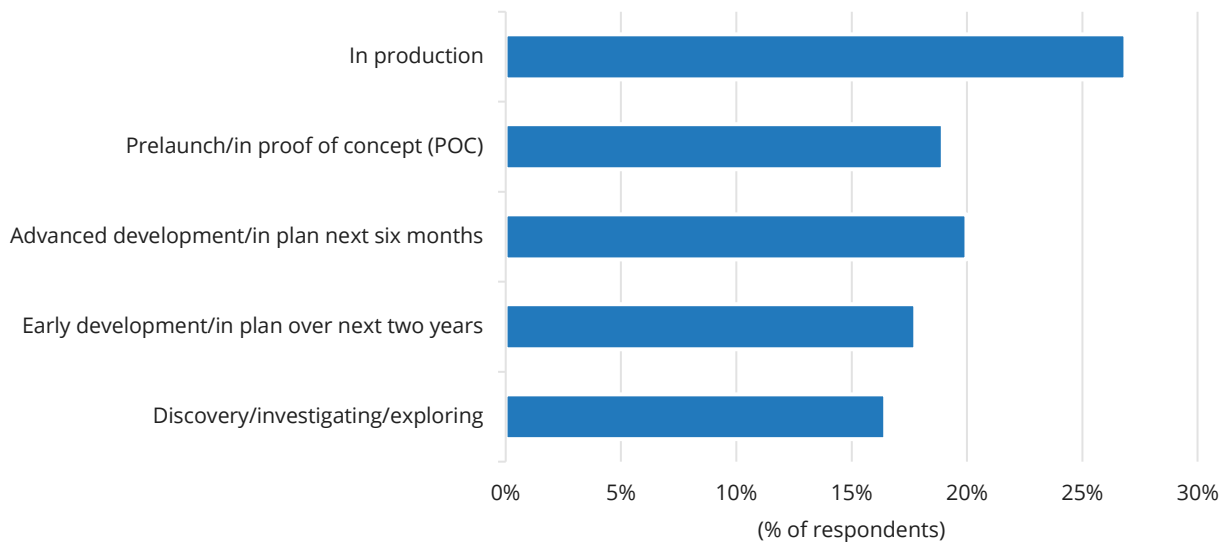Note: Multiple responses were allowed.

Source: IDC's *ML Platforms BuyerView Survey,* February 2022

While global adoption of AI/ML technologies is increasing, customers are still in the initial stages of AI/ML adoption. In IDC's February 2022 *ML Platforms BuyerView Survey*,[1] only about 27% of respondents indicated that they have AI/ML initiatives running in production (see Figure 2). Most customers also reported that they have faced failures in their AI/ML initiatives. Some of the primary reasons for those failures are the lack of production-ready data pipelines, upskilling/staffing efforts, and the need to reengineer applications.

FIGURE 2

AI/ML Adoption Maturity

Q.   *What percentage of your organization's AI/ML initiatives are in each of the following stages? Please provide your best estimates.*

n = 311

Base = all respondents

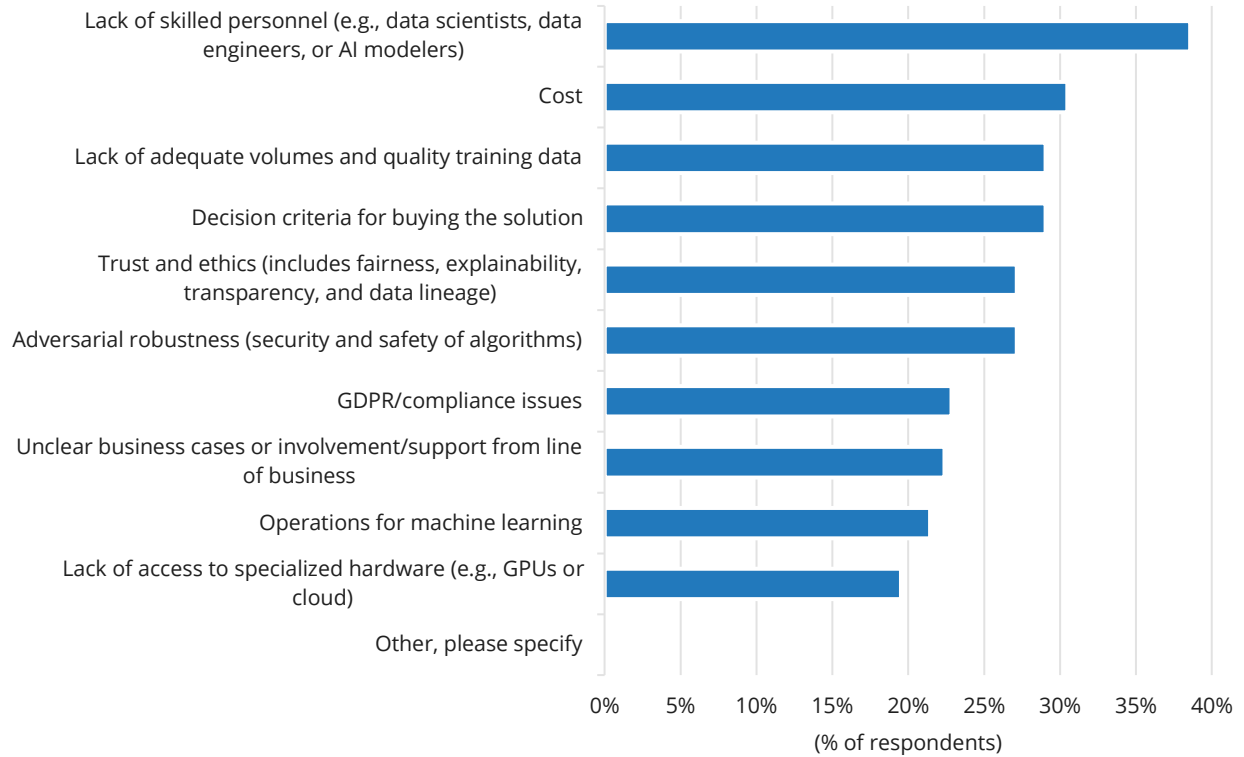Source: IDC's *ML Platforms BuyerView Survey,* February 2022

---

[1] All references to the accompanying survey refer to this IDC survey, unless otherwise specified.

Respondents also cited lack of expertise and cost as some of the main challenges they face in implementing AI/ML solutions (see Figure 3).

FIGURE 3

AI/ML Adoption Challenges

Q. What are the main challenges with implementing AI/ML technology at your organization?

n = 210

Base = Persona 1 respondents

Note: Multiple responses were allowed.

Source: IDC's *ML Platforms BuyerView Survey,* February 2022

The AI/ML life cycle involves multiple stages including data preparation, model build and training, model evaluation, model deployment, and model monitoring. Customers have a choice of leveraging multiple platforms, languages, and libraries for implementing these stages. IDC observes consolidation of tools and software around model build/training platforms/tools/languages with clear winners standing apart. However, IDC also observes that customers use diverse tools and software for model deployment/MLOps needs without any standardization yet.

So, What Defines a True Enterprise-Grade AI/ML Platform?

Leaders evaluating the myriad available options need to ensure that optimizing for today does not prevent them from flexibly meeting their needs tomorrow.

With the challenge and complexity of disparate solutions, a true enterprise-grade AI/ML platform provides interoperability, scalable provisioning of services, and a shared unified experience to complete any task from data ingestion to model building to model deployment and governance.

This white paper presents the top questions you need to consider when choosing the right data and AI/ML platform for your organization. This is a crucial decision; the right platform can help you improve business outcomes, drive innovation and digital transformation, and achieve successful ROI.

> **IDC expects the MLOps ecosystem to mature in the next 12–18 months, with platforms that support the AI/ML life cycle end to end, from data ingest to model monitoring with unified governance gaining a larger market share**.

PRACTICAL CONSIDERATIONS WHILE CHOOSING AN ENTERPRISE-GRADE AI/ML PLATFORM

1. Does the Platform Support Data Ingest and Management from Multiple Clouds, SaaS Applications, and Services?

If you are leveraging AI/ML capabilities for innovative and complex use cases, you are most likely to consume heterogeneous data sources and data types. You may also be deploying machine learning models across multiple deployment environments – public cloud, on premises, and edge locations. In IDC's February 2022 *ML Platforms BuyerView Survey,* respondents indicated usage of enterprise data, objects, and transactional and streaming data. About one-fourth of the respondents also indicated building and deploying models across multiple deployment locations. With such diversity in model data types, it is important that the ML platform support data from multiple cloud platforms, SaaS applications, and services.

IDC recommends that organizations selecting an ML platform for their machine learning needs consider the questions on data management in the sections that follow.

What Is the Distribution of Your Business Applications Among SaaS, Cloud Applications (Such as Salesforce, Workday, and SAP), and Modern Applications?

Enterprises are modernizing enterprise workloads as part of the digital transformation journeys they are on. They are taking one or more of the following approaches toward application modernization:

- Adopting equivalent SaaS applications
- Migrating workloads to cloud platforms
- Re-architecting to modern/cloud-native architectures
- Implementing net-new, modern applications (such as AI/ML workloads)

As you embark on your digital transformation journeys, you may also be modernizing your business applications through one or more of the approaches mentioned previously. The distribution of the applications you are using (SaaS, cloud, and modern) depends on the stage of your digital transformation, the characteristics of your workloads, and your preferences. This distribution has an enormous influence on your choice of data platforms and how you consume AI capabilities.

Does Your Platform Support Multiple Cloud Environments?

Customers choose multiple cloud platforms, each for their own specialized capability. Respondents to the survey cited integrated data management capabilities as the top reason for selecting AWS and IBM, security and compliance considerations for selecting Azure and GCP, and streamlined deployment for selecting Oracle Cloud. Leveraging multiple cloud platforms also minimizes the dependency on a single provider and improves data resiliency. If you are not currently using multiple cloud environments, it is just a matter of time before you start. The platform you invest in needs to support multiple cloud environments.

Do You Concurrently Use Separate Cloud Providers for Multiple or Single Workloads?

Although you may be employing multiple cloud platforms, you are most likely treating these cloud environments as siloed environments — often one cloud platform as the primary environment and the other as a backup environment. Since cloud platforms are not easily interoperable, it is difficult to deploy the same application concurrently across multiple cloud platforms. If you ever need to do so for your ML workloads, it is important that the machine learning platform support such concurrent use of different cloud service providers.

Do You Have Resiliency, Security, and Governance Across Geographies Covered?

Line-of-business (LOB) applications need to be resilient and highly available. These applications also need to ensure the security of user data and user privacy and comply with applicable regulations. As you scale your applications to multiple geographic locations, you need to ensure the applications are secure, resilient, and compliant across all the geographies covered. The ML platform should support enabling these attributes across all geographies.
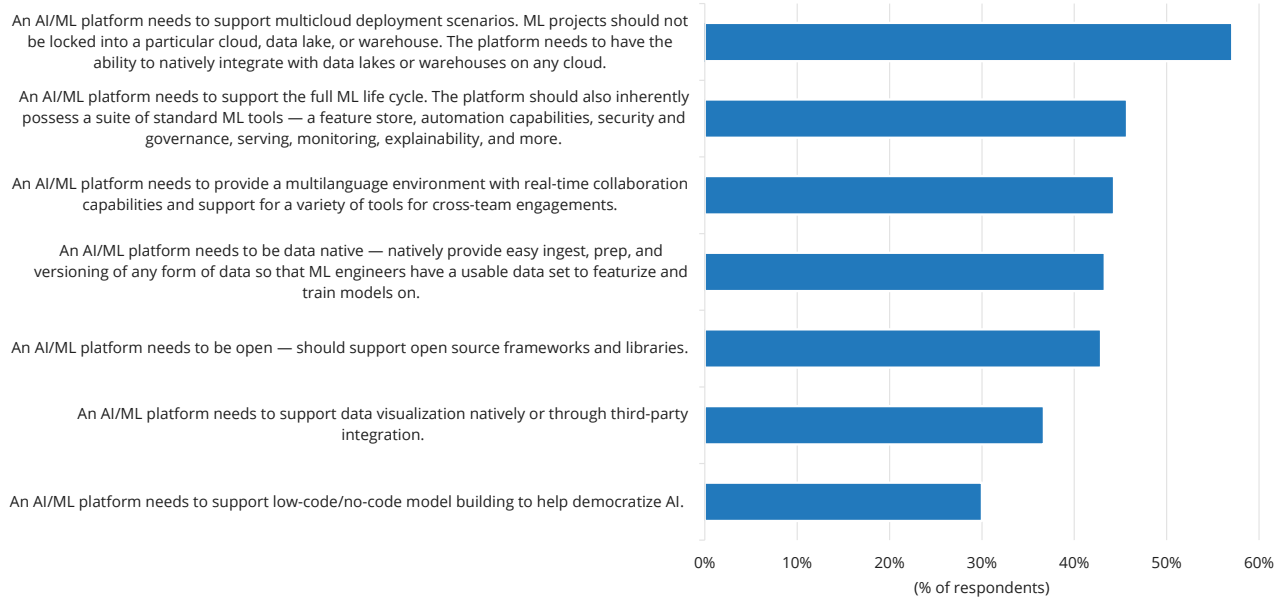
2. Does the Platform Support End-to-End AI/ML Life-Cycle Management?

Machine learning models go through various stages across AI/ML life-cycle management — data ingestion, data preparation, model build, model evaluation, model deployment, and model monitoring. Each of these steps involves different personas, including DataOps engineers, data scientists, ML engineers, and IT/DevOps engineers. Often, end users may need multiple tools and platforms for these stages. A machine learning platform that supports all these stages end to end can significantly reduce the complexity of using multiple frameworks. Forty-six percent of survey respondents indicated that an AI/ML platform needs to support the full machine learning life cycle, making it one of the top 3 AI/ML platform capabilities sought (see Figure 4).

FIGURE 4

Key AI/ML Platform Capabilities Sought

Q. What do you think are the critical capabilities of an AI/ML platform?

An AI/ML platform needs to support multicloud deployment scenarios. ML projects should not be locked into a particular cloud, data lake, or warehouse. The platform needs to have the ability to natively integrate with data lakes or warehouses on any cloud.

An AI/ML platform needs to support the full ML life cycle. The platform should also inherently possess a suite of standard ML tools — a feature store, automation capabilities, security and governance, serving, monitoring, explainability, and more.

An AI/ML platform needs to provide a multilanguage environment with real-time collaboration capabilities and support for a variety of tools for cross-team engagements.

An AI/ML platform needs to be data native — natively provide easy ingest, prep, and versioning of any form of data so that ML engineers have a usable data set to featurize and train models on.

An AI/ML platform needs to be open — should support open source frameworks and libraries.

An AI/ML platform needs to support data visualization natively or through third-party integration.

An AI/ML platform needs to support low-code/no-code model building to help democratize AI.

(% of respondents)

n = 210

Base = Persona 1 respondents

Note: Multiple responses were allowed.

Source: IDC's *ML Platforms BuyerView Survey,* February 2022

IDC recommends that organizations selecting an ML platform for their machine learning needs consider the questions on AI/ML life-cycle management in the sections that follow.

Are You Able to Automate Governance Policies to Secure, Catalog, and Share Data of All Sources and Data Structures So You Can Comply with Regulations as well as Reduce Manual Work?

With no universal regulations in place for ML data, it is up to you to define what kind of regulations you need to enforce. For you to successfully enforce governance regulations, the ML platform you employ needs to be able to support automated governance policies and to enable end users to secure, catalog, and share data.

Does the Platform Support Multiple Data Sources (Such as Structured, Unstructured, and Streaming) and Multiple Data Management Platforms (RDBMS, Data Lake, Object Store, etc.), Thereby Proving to Be a One-Stop Solution Instead of Having to Orchestrate Multiple Services Across Vendors?

You are often leveraging multiple data types and sources for machine learning needs. For example, it is common to employ data management platforms to store user data, relational databases for transactional data, object stores for multimedia files, data warehouses for enterprise data, and streaming platforms for streaming data. With such varied data types and platforms in the mix, it is important that the ML platform you use provide native support for various data and AI solutions. Otherwise, you may end up having to orchestrate multiple services across vendors through custom code or additional work.

3. Is the Platform Open and Flexible?

End users prefer flexibility and choice with the tools and software they employ for data management, model build/training, model inferencing, programming languages/libraries, and model operations. Such flexibility enables the end user to employ best-of-breed tools for the task at hand. It also helps build upon existing investments in tools and software. About 44% of respondents cited support for multilanguage environments with real-time collaboration as one of their top 3 requirements from an AI/ML platform (refer back to Figure 4).

IDC recommends that organizations selecting an ML platform for their machine learning needs consider the questions on flexibility and choice in the sections that follow.

Does the Platform Support Choice of Leading Frameworks Like TensorFlow, PyTorch, Apache MXNet, Microsoft CNTK, and Scikit-Learn; Development Environments Like Jupyter Notebook; and Languages Like Python and Java?

As you expand upon your machine learning use cases, you may start using a disparate set of tools, languages, libraries, and frameworks based on your needs. The accompanying survey shows that end users employ multiple frameworks such as TensorFlow, PyTorch, Caffe, and scikit-learn (in that order) for their ML/DL needs. End users also employ diverse programming languages such as Python, Java, and R and development environments such as Jupyter Notebook, Visual Studio, and cloud-based integrated development environments (IDEs) to develop machine learning models.

With such a variety of tools, libraries, languages, and frameworks in the mix, it is important that the ML platform provide choice and flexibility with these aspects. This flexibility enables you to use the tools, libraries, languages, and frameworks of your choice for your machine learning needs.

Does the Platform Support Collaboration Among Data Scientists, Analysts, Machine Learning Engineers, and Legal and Compliance Functions?

As discussed previously, a machine learning model goes through multiple stages throughout its life cycle. As a model goes through these stages in your machine learning pipeline, it is managed and monitored by multiple personas like the data scientists, business analysts, data and machine learning engineers, and legal and compliance officers from your organization. Typically, these personas work in their silos with their own set of tooling, which makes it difficult to collaborate. To improve your overall employee productivity and model velocity, the ML platform you use needs to support and increase collaboration among these personas in your organization.

Does the Vendor Participate in Open Source Efforts? Does the Platform Provide an Easy Ramp-Up for Fast-Changing Open Source Innovations?

Open source development has accelerated innovations across technologies such as cloud computing, platform as a service (PaaS), and AI/ML rapidly. The average life cycle of technology innovations has shrunk from multiple decades to less than a decade. The pace of adoption of such innovations has also increased rapidly. For example, while it took more than a decade for cloud service providers to cross $100 billion in cloud services revenue, AI software vendors hit that mark much faster.

While open source development accelerates product innovations, faster development cycles may also bring instability to the product. It is important that the vendor understand open source development and participate in open source efforts so that it can provide the product with much-needed stability and robustness.

Along with such rapid innovation, the decision-making process also becomes complex. Often, enterprise customers tend to become perplexed with the options available, and you may be too. The platform needs to support easy onboarding and ramp-up for these innovations so that you, as an end user, can build up on your success with adopting these innovations.

4. Is the Platform Easy to Use?

Lack of expertise is repeatedly cited as one of the inhibitors to AI adoption. Data science, ML development, and ML operations all have a steep learning curve. According to IDC's 2022 *StrategiesView Survey,* 53% of respondents reported a lack of skilled talent as one of the top 5 challenges in scaling their AI initiatives. With the increasing complexity of ML use cases, expanding matrix of data types/sources, and rapidly growing technology innovations, end users want easy-to-use ML platforms. They also want the platform to provide self-service and low-code/no-code model building capabilities to help them overcome the challenges of lack of expertise.

IDC recommends that organizations selecting an ML platform for their machine learning needs consider the questions on ease of use in the sections that follow.

Does the Platform Support Self-Service? Are All the Data and ML Services Interoperable, and Do They Enable Self-Service Across the Organization?

Self-service capabilities help data scientists/business analysts select the right ML algorithms, build ML models, and provision the required infrastructure for training with less dependence on IT. Such independence helps them build and iterate on models faster, thereby enabling them to build AI capabilities faster. To enable you to build and iterate on your machine learning models faster, the ML platform you invest in needs to support such a self-service-centric way of model build and delivery.

Does the Platform Support Prebuilt ML Models/Integration of Third-Party Recipes?

There will be instances when you want to craft custom AI models to address specific outcomes. However, if your platform has predefined AI workflows and industry solutions built in, you can automate complex processes and optimize your employees' time instead of building each model from the ground up. Prebuilt ML models enable data scientists to get started with their models and use cases faster. There is also an increase in behaviors such as sharing one's recipes with peers and consuming third-party recipes in the spirit of community. Data scientists can also tweak such models/recipes to suit their needs, thus providing them with more control than low-code/no-code approaches. If you are a data scientist looking for a quick start with building machine learning models, use an ML platform that supports

such prebuilt ML models and enables integration with third-party recipes. Look for a platform that provides pretrained AI models that help you quickly tap into organizational data and improve efficiencies in areas such as finance, business operations, customer care, and supply chains.

Does the Platform Have Tools to Help Your Team Members, Regardless of Their Data Science Skill Level, Automate the Entire ML Life Cycle?

As you scale your machine learning pipelines, you are most likely to meet challenges with lack of automation across the entire ML life cycle. Along with lack of expertise, lack of automation is one of the top challenges for AI adoption. According to IDC's 2022 *StrategiesView Survey,* 52% of respondents reported a lack of automation/MLOps as one of their top challenges to scaling AI initiatives. To help your team members, irrespective of their data science expertise, your ML platform needs to have the right tooling and capabilities to automate the entire ML life cycle.

Does the Platform Provide a Simple, Intuitive, and Easy-to-Use Interface?

Apart from self-service capabilities, prebuilt ML algorithms, and end-to-end automation, your ML platform also needs to provide a simple, intuitive, and easy-to-use interface. Such an interface helps not only the data scientists but also all other personas, including ML engineers, IT operators, and LOB owners.

Does the Platform Support Low-Code/No-Code ML or Auto-ML Capabilities?

Low-code/no-code model building capabilities allow personas with less data science experience, such as LOB owners and CXOs, to build AI-enabled applications faster. Auto-ML capabilities for building models automatically from data sets provide a great starting point for data scientists with a complete model that can be tweaked according to their needs. If you are an LOB owner or a C-suite executive interested in building AI-enabled applications quickly, your ML platform needs to be able to provide these low-code/no-code or auto-ML capabilities for building machine learning models.

5. Does the Platform Enable Responsible AI Initiatives in Safe and Secure Ways?

AI can have a positive impact on our lives, but we are also witnessing increased occurrences of bias/trust incidents. Bias, intentional or not, can result in the exclusion of certain groups of people, unfair decisions, unexpected behavior from AI systems, unethical outcomes, or even fatalities. Such incidents have resulted in reduced public trust, loss of revenue, damage to brand reputation, and criminal investigations. As such, AI systems should be designed, developed, and deployed responsibly so that the decisions made are ethical, trustworthy, accountable, safe, and secure.

IDC recommends that organizations selecting an ML platform for their machine learning needs consider the questions relevant to responsible AI implementations in the sections that follow.

Does the Platform Support Secure and Safe Handling of Machine Learning Data and Support Data Governance, Including a Shared Metadata Catalog?

Along with the increased adoption of AI capabilities, concerns about security and user privacy are also growing. Organizations and governments across the world are taking measures to ensure that AI capabilities are used without breaching user privacy and protecting user data. To address these concerns, the ML platform needs to support the secure and safe handling of ML data. It also needs to provide methods to enforce data governance, encrypt data at rest, and share metadata securely. If you care about securing user data and ensuring user privacy (which you must), you need to make sure your ML platform supports secure and safe handling of ML data and enforces data governance.

Can the Platform Accurately Explain AI Model Outcomes and Safeguard Fairness by Minimizing Model Bias and Drift?

Explainability refers to the ability to justify the decisions and conclusions made by the AI system. While data scientists and auditors can explain model decisions, decision makers may need such explainability to quickly approve/disapprove ML models.

AI systems should be fair and unbiased to avoid any unintentional unfair treatment of certain groups. Model bias could creep from multiple sources – training data as well as incorrect model build, selection, or deployment. AI systems should use training data and models that are free of bias and should have correct checks and balances to ensure that the system does not discriminate based on gender, race, color, orientation, faith, or anything else.

To enable you to explain model decisions/outcomes better and ensure fairness by minimizing model bias and drift, your ML platform should provide for model explainability and be able to detect, notify, and help troubleshoot model drift.

Does the Platform Provide for Model Governance?

Model governance enforces consistent policies through the AI life cycle, enables a centralized view across all stages, improves collaboration among all personas, and enhances model accuracy/user experience. Without universal regulations and standards for implementing responsible AI, it is important that the ML platform provide for model governance so that you, as an end user, can define what responsible AI means for you and then execute on the same.

CHOOSING A PROVIDER

Successful AI adoption hinges on more than just platform capabilities. As you progress through your AI adoption journey, you may have varied needs. At the beginning of your journey, you may need more help with vendor and technology selection, building prototypes, and training. When you are well into the adoption journey, you need support for scaling, in-depth training, and better pricing models.

IDC recommends considering sales, pricing, and support strategies along with platform capabilities when selecting your provider.
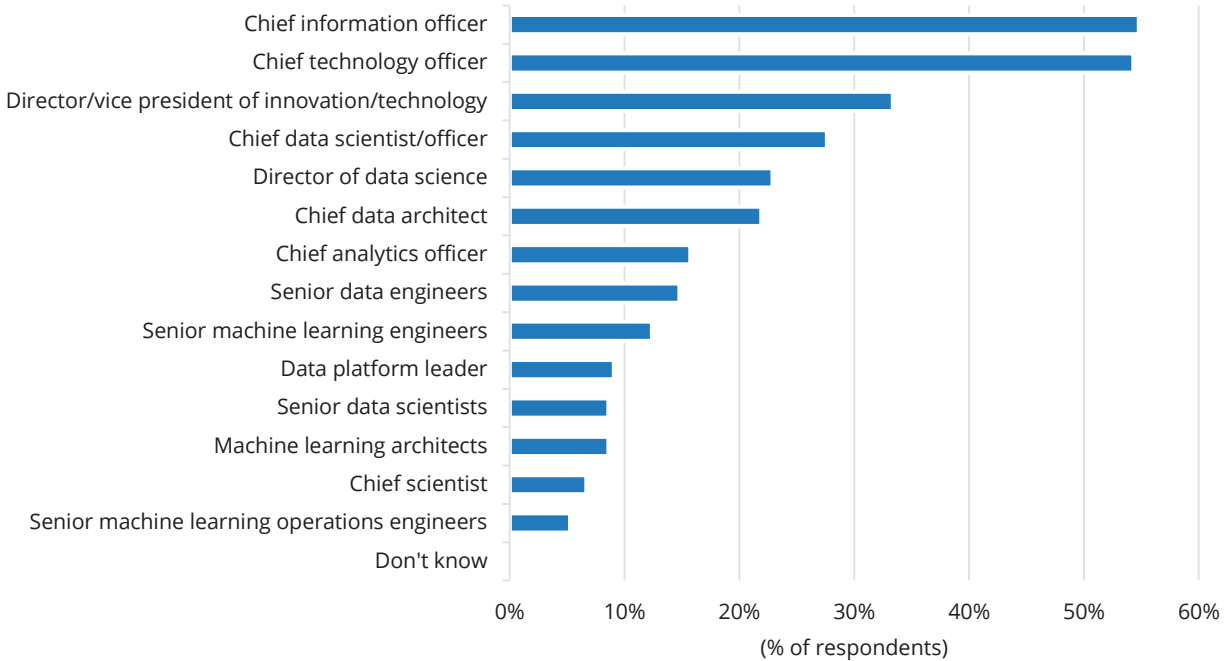
Sales Strategy

If you are like most enterprise organizations, your AI purchasing decisions are most likely influenced and made by your C-suite. About 55% of survey respondents indicated that the CIO/CTO is responsible for AI/ML platform selection, while roughly 33% of survey respondents indicated the director/vice president is responsible (see Figure 5).

FIGURE 5

AI/ML Platform Decision Makers

Q.   Which of the following roles are responsible for AI/ML platform selection?

n = 210

Base = Persona 1 respondents

Note: Multiple responses were allowed.

Source: IDC's *ML Platforms BuyerView Survey,* February 2022

If you are a CXO, you may need to pay attention to the vendor's sales strategy to evaluate the vendor's ability to cater to both the C-suite and the practitioners – a successful vendor would take a two-pronged sale strategy to appeal to both the C-suite (top down, enterprise sales approach) and the practitioner (bottom up, driven by the user/community through nonconventional channels such as Slack). Such a sales strategy indicates that the vendor is better positioned to support your needs throughout the AI life cycle.

Pricing Strategy

As you mature in your AI adoption, you may soon notice that your AI spend is spiraling. With cost being one of the topmost concerns with AI adoption, you may be looking for more control of your AI spend. AI/ML platform vendors usually provide four types of pricing options:

- Fixed commercial licensing
- Usage-based pay-as-you-go pricing
- Flexible spending
- Reserved transactions

Each option may provide more variations such as compute or storage use for the pay-as-you-go pricing model. About 26% of the respondents to the accompanying survey indicated a preference for flexible spending with an ability to spend across all services up to the prepurchased amount at a discount. Some vendors also offer fine granularity with the pay-as-you-go model (per-second billing). Such options may prove to be cost effective during model training and inferencing. You need to select the right pricing model for your needs, fully leveraging the levers and granularities your ML platform provides.

Support Strategy

Your support needs may vary depending on your AI adoption maturity and your in-house expertise. As you progress in your AI adoption maturity, you need to be supported through POC, training, buildout, certification, and more. Select the partner that can support your needs, not just during your current stage but throughout your AI adoption journey.

SUMMARY AND CALL TO ACTION

Today, enterprises are confronted with a complicated set of business challenges, including an increasing pace of business, an expanding volume of business data, the need to think about shared data strategies to truly derive value from data, a growing scope of global commerce, and a multitude of risks for customers, employees, and suppliers. The volume of customers and suppliers, along with regulatory complexity and multi-industry businesses, means complexity is common in global businesses. Disruption is unsettling, but it can also serve as a catalyst for innovation and transformation. We have now entered the domain of AI-augmented work and decision making across all the functional areas of a business across industries.

Selecting the right ML platform and provider makes all the difference in your success with AI adoption. The right platform can help you improve business outcomes, drive innovation and digital transformation, and achieve successful ROI. The right provider can assist you at each stage of this journey by providing support, training, advisory services, and pricing options.

IDC recommends asking the following five questions when selecting the right ML platform/provider and choosing wisely:

- Does the platform support data ingest and management from multiple clouds, SaaS applications, and services?
- Does the platform support end-to-end AI/ML life-cycle management?
- Is the platform open and flexible?
- Is the platform easy to use?
- Does the platform enable responsible AI initiatives in safe and secure ways?

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com