Verta

VERTAINSIGHTS RESEARCH

The State of Machine Learning Operations

2022

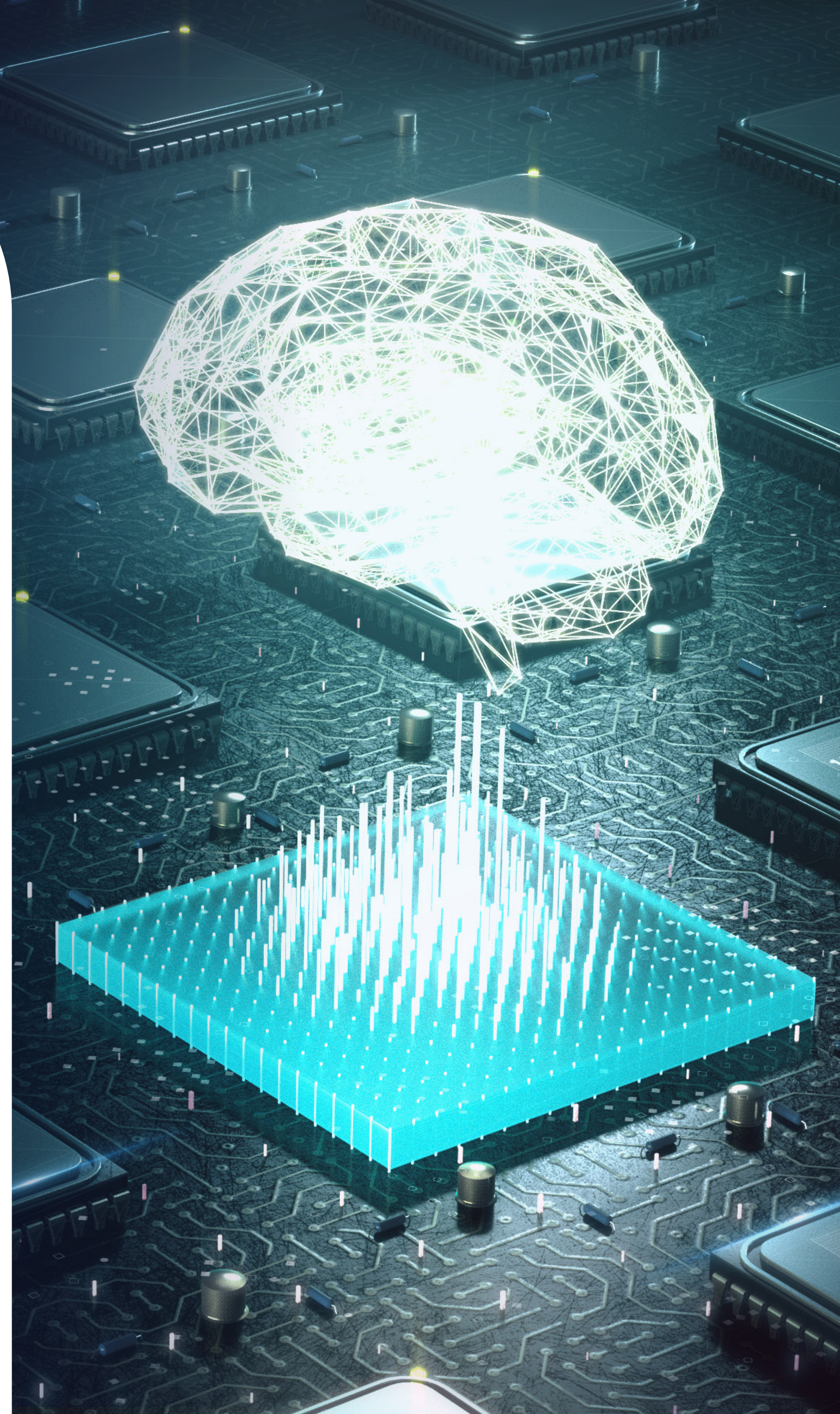


Table of Contents

The AI/ML Paradigm Has Shifted

Introduction from Manasi Vartak

Artificial intelligence is transforming how consumers experience technology, whether through smartphones in our hands, smart devices in our homes, or the intelligent applications and services that we interact with every day.

The consequences of this transformation are enormous for businesses. AI's growing impact is forcing organizations to rethink risks and opportunities, the investments necessary to become AI-forward, and the potential to disrupt industries and become leaders — or be disrupted and become laggards.

The State of Machine Learning Operations study from Verta Insights set out to explore trends, best practices, challenges, and opportunities shaping the industry — from model experimentation and deployment through real-time safe operations. Importantly, the study revealed differences in how financially successful organizations approach ML versus laggard peers.

A key finding of the study is that while organizations have made significant investments in their data science teams and the tooling necessary to enable model building and training, they still struggle to operationalize their models and manage their models in production. In other words, **companies continue to be challenged to realize value from their investments in Build and Train.**

It's not surprising, then, that we are starting to see a paradigm shift in how organizations are prioritizing their investments around AI/ML. Executives are putting more focus on downstream ML capabilities, like monitoring, features stores, and model catalogs. The implication is that the ability to build complex, interesting models is no longer the competitive differentiator it once was - a startup can combine some pre-trained models with a great user interface and disrupt an entire industry.

Instead, competitive differentiation will now be a product of an organization's ability to effectively and efficiently operationalize ML. The paradigm shift that we see evidenced in this study is the movement away from a focus on data science and toward a mindset where we treat models like software. In adopting an Operational AI mindset, leaders are ahead of the game in delivering better performance — and therefore better value — from their AI/ML teams. Although, as we'll see in the study, even leaders have a long way to go before they achieve the kind of "five 9s" performance from AI/ML that we expect from other business-critical enterprise applications.

I want to thank all the study participants for generously sharing their time and thoughts on the State of MLOps. We established Verta Insights to conduct groundbreaking research and deliver insights that assist AI/ML practitioners and executive leaders in preparing their organizations for the AI-enabled future. This study advances that goal and, we hope, will assist our industry in realizing the full potential of AI and ML.

Manasi Vartak
CEO & Founder, Verta

Summary of Key Findings

01

ML Models Are Business Critical

Companies see models as a key driver for their success: 86% of participants in the study said that models are significant to extremely significant for their company's success. What's more, companies are building a large number of models: More than half the participants reported that they have more than 50 models in one stage of development or production.

02

Operational ML is Growing Rapidly

Real-time or low-latency use cases already predominate among the study participants, who report that 54% of their use cases already fall into this category (versus batch use cases). But that percentage is set to increase markedly over the next three years, as 69% of participants said that they expect real-time/low-latency use cases to increase (44%) or increase significantly (25%) in that time period.

03

ML Investments Now Targeting Downstream Capabilities

Business investment priorities are shifting from initial model building tools and infrastructure to downstream operations. Many organizations appear mature in their tooling and infrastructure for fundamental model building and training — on average, companies report using 6 different tools for model build and training. These organizations are now focusing on implementing more downstream operational AI capabilities like monitoring, feature stores, and model catalogs.

04

Financially Successful Companies Outshine Peers in ML Performance

Leaders are organizations that outperform their peers financially, and in our study companies categorized as leaders also outperformed their peers in several areas of business performance related to machine learning. Leaders are more than 2x likely than laggard peers to successfully ship AI projects, products or features, and 3x more likely to meet service level agreements for uptime performance in operations.

05

Leaders Are Making Specific Investments to Support Operational AI

Organizations that outperform their peers financially have made specific AI/ML investments at higher rates than their laggard peers. In particular, leaders have invested in platform teams and MLOps platforms at a higher level than peers. And nearly twice as many leaders as laggards have invested in creating a Model Governance or Ethics Council that provides governance around the risks and usage of AI.

06

Despite Advances in MLOps Practices, ML Performance Still Lags

Looking at MLOps relative to other technologies or software development, it's painfully clear that despite leaders outperforming their peers, MLOps remains far from the performance and reliability that we expect to see in other business critical applications or systems. Just 1 in 5 organizations report having at least 80% of their models operating reliably in production. That's far from the "five 9s" that we would expect in terms of performance in other critical enterprise systems.

Study Findings

ML Models Are Business Critical

Despite the fact that we are still in the so-called “early innings” of AI, it’s now taken for granted that models are essential to business success. Participants in the State of MLOps study overwhelmingly said that models are significant to extremely significant to their companies’ revenue-generating or customer-facing products and services.

The study also asked participants to estimate how many models their company has developed, and a little more than half reported that they have more than 50 models in one stage of development or production - including 18% that reported having more than 500 models. (See charts on next page.)

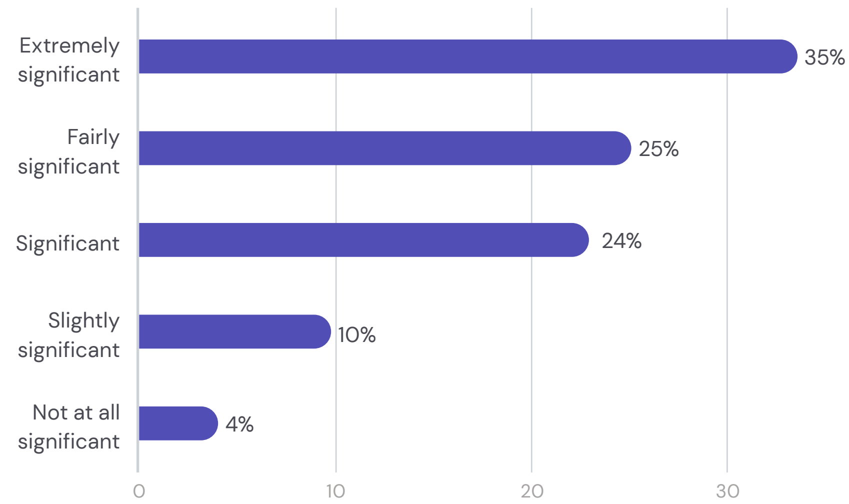
A follow-up question revealed that oftentimes participants were aware of only the number of models produced within their team, department or business unit, without visibility into ML activity elsewhere in their organization.

These results do beg the question, “What is a model?” For example, if a company has a model running on four different devices, is that a single model, or is it four models? Do we count models that have been archived, even if we need to keep track of those models for governance and audit purposes? Currently the industry does not have a standard around this.

Notably, elsewhere in the study we see that many companies are re-training their models on at least a monthly basis. Simple math suggests that, if we multiply 50 or 100 models in production by 12 updates a year, companies could find themselves trying to manage 600, 1000, or more models or model versions.

Importance of Models

How significant are models to your company’s revenue-generating or customer-facing products and services?



N=208

> “Companies are transitioning from a handful of models to a fleet of models. As they reach that inflection point, organizations need to transition from artisanal, one-off approaches for managing models to an industrialized approach with the appropriate tooling and standardized processes.”

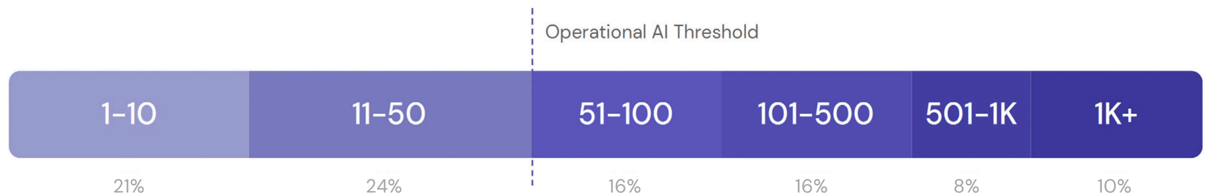
Meeta Dash
Vice President of Product, Verta

The sheer volume of models that participants report they are managing is significant and suggests that AI/ML is building a great deal of traction. This raises two considerations for companies advancing on their AI/ML journey:

- As the number of models increases, what are the implications for technical debt that we are bringing into our ML operations?
- And, as our portfolio of models expands, how do we need to change our tools and processes for managing models?

The implication is that as companies increase the number of models they are working with, and as models become more critical to their business success, organizations must consider putting in place standardized tooling and systematized processes for managing their models now to ensure they are getting the optimal value out of their investments in data science.

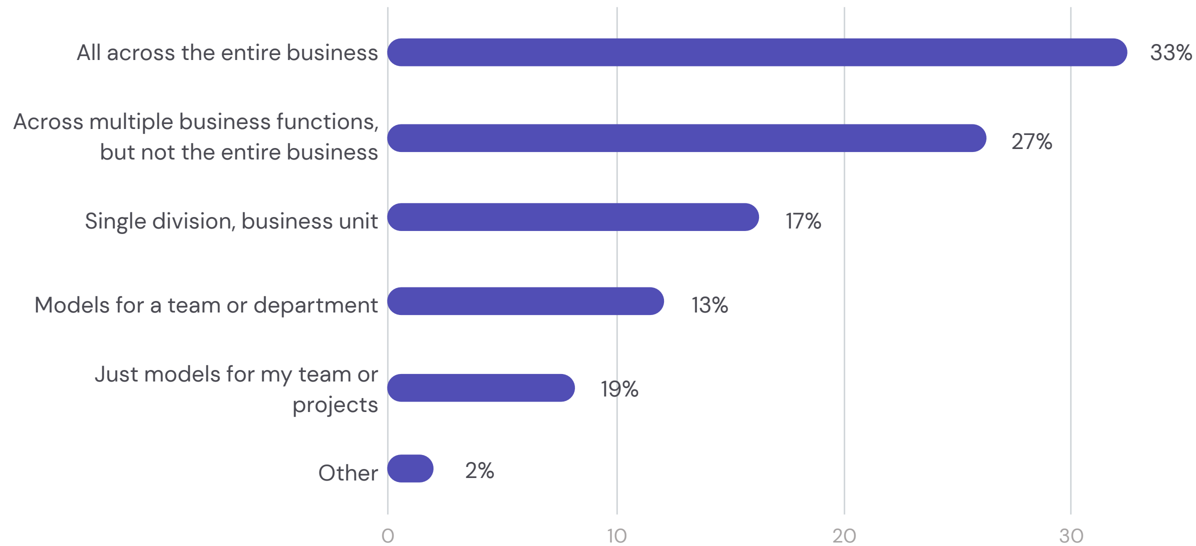
Approximately, how many total models has your company or organization developed?



N=208

Number of Models

The number of models you cited in the previous question represents:



N=208

Investment Priorities

In the study, we asked about business investment priorities to understand where organizations are focusing their efforts to enable and improve the machine learning lifecycle. Participants could select all the different areas where their companies are investing, and the result is a rank-stacking of investment priorities. This provides a picture of the building blocks that companies are putting in place as they are currently working toward their vision for machine learning.

To an extent, of course, these priorities reflect the different maturity levels of various technology segments and where companies are in their ML journey. For example:

ML monitoring

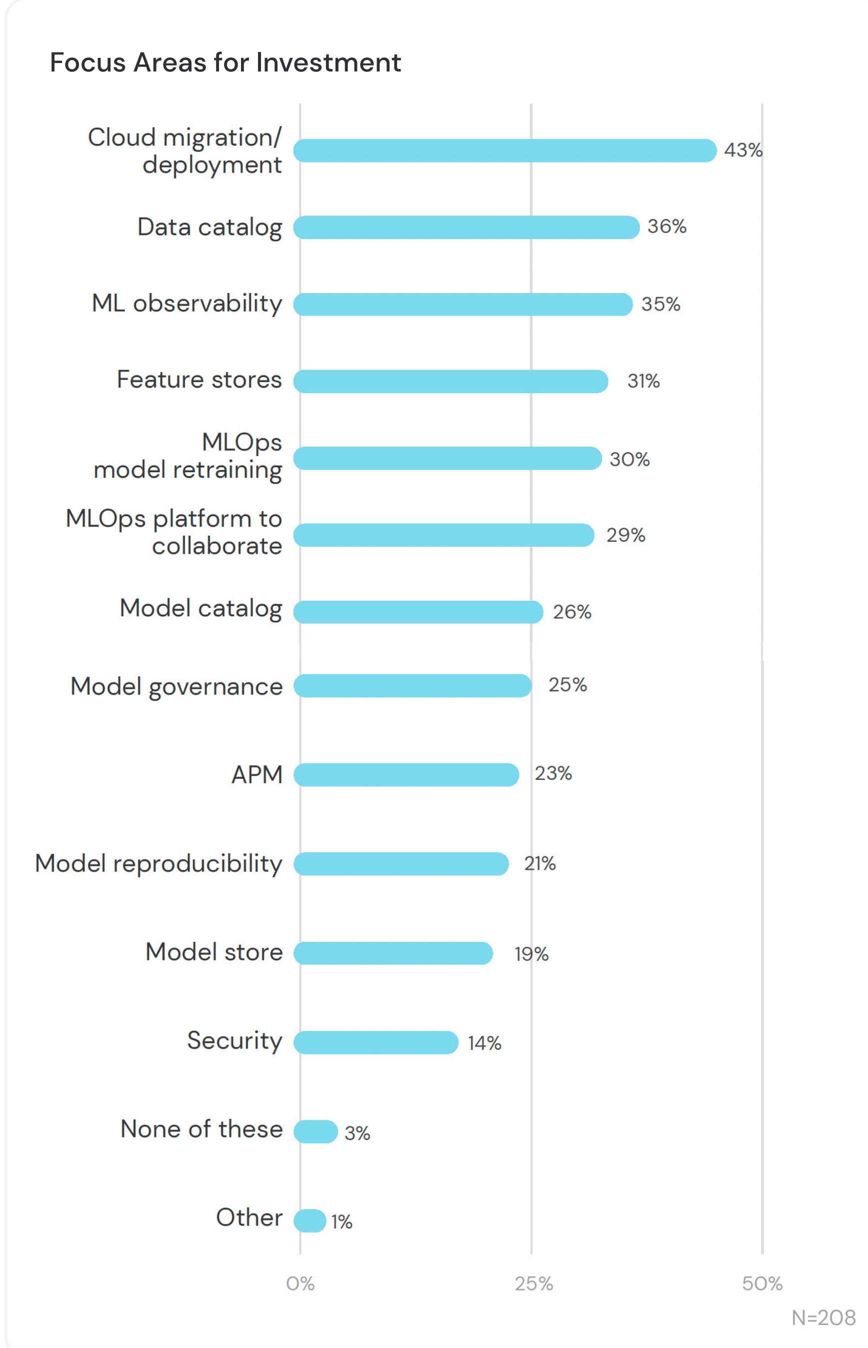
A company currently investing in monitoring likely has already made investments in model building and training, and they are now focused on monitoring their models in production.

Feature stores

Organizations begin investing in feature stores typically when their ML operations are advancing beyond batch and they are increasing their use of real-time use cases.

Governance

Once an organization has matured to a scale where they can no longer manage their model portfolio with manual processes and spreadsheets, they need to put in place tools like a model catalog to enable model governance and management.



Investments Made to Date

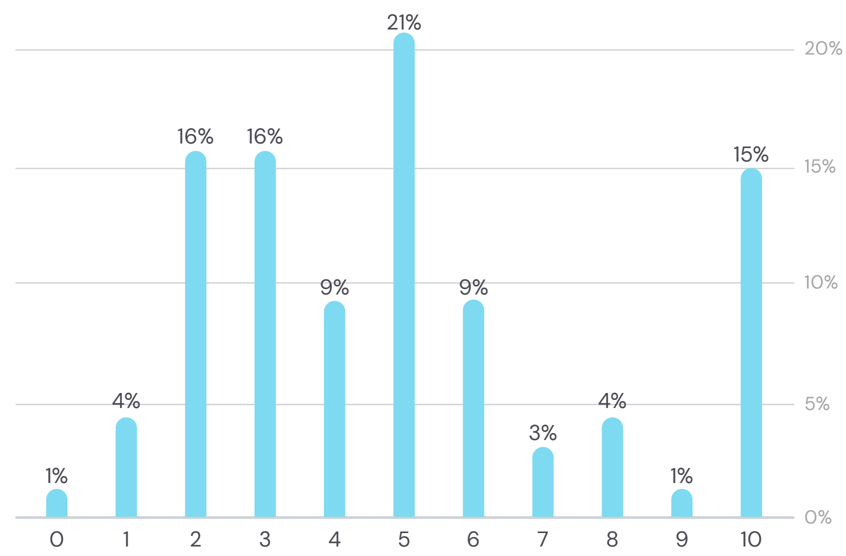
While the focus of the prior section was around where companies currently are putting their investments, the study also asked participants about the technologies that their companies already have invested in to date. Of course, companies can invest in different ways, including in internal development to build their own tools (i.e., do it yourself), or in external enabling technologies. The responses to this question provide a picture of the building blocks that our study participants say their companies have already put in place and are using today to enable machine learning operations at their organizations.

Key Highlights

- On average, the participants said their companies are using six different tools for ML
- 15% of participants said that their companies are using 10 or more different tools
- In total, participants cited more than 40 different tools

The large numbers of tools cited speak to the fragmentation of this market: Companies are using a host of different technologies to support machine learning. This fragmentation presents challenges, of course, to the vast majority of organizations that cannot afford to build their own infrastructure. These companies must stitch together a set of external tools to support their ML operations or, alternatively, deploy a platform capable of performing the functions of several different technologies while also integrating with external tools. Taking the latter, platform-centric approach can minimize the change management overhead associated with deploying multiple tools across disparate teams, and reduce the burden of tracking product roadmaps across a large number of tool providers.

Tools Currently in Use



N=139

> "For organizations that already have tools in place to support their data science teams with model building and training, it makes sense that they are now shifting their investments to focus on tooling to manage and monitor their models in production—which, after all, is where models truly deliver value."

Rory King
Head of Verta Insights

Operational ML is Growing Rapidly

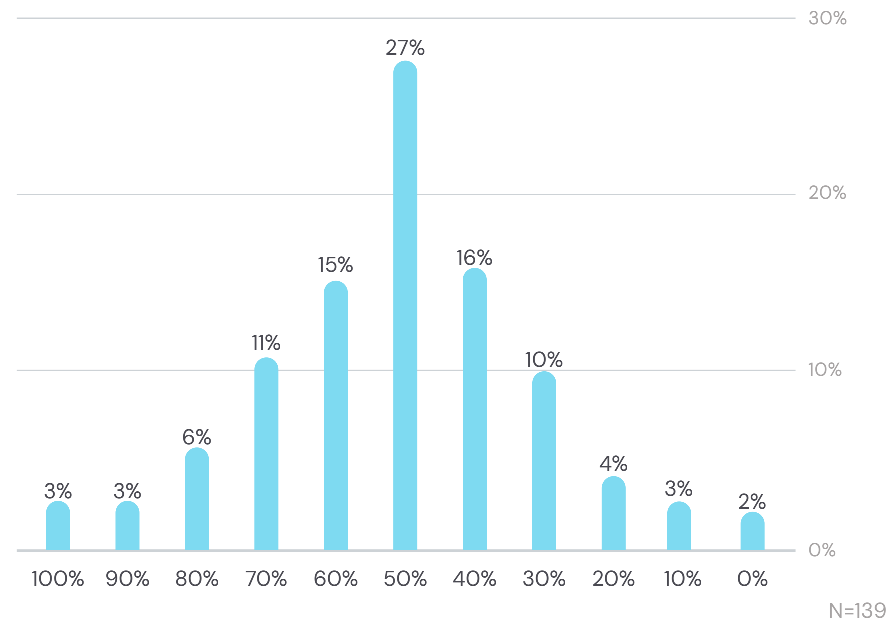
The study asked participants about the balance between analytical (i.e., batch) use cases, versus operational AI, or real-time/low-latency, use cases employed at their organizations.

To level-set, Operational AI refers to models that are in use in intelligent applications, with some level of real-time inference, and continually operating to a high standard. Real-time implies a very rapid response - think less than 100 milliseconds - often because a customer depends on getting a response very quickly.

Study participants reported that just over half (54%) of their models are running in real-time or low-latency use cases or applications currently. However, just over two-thirds (69%) of our participants said that they expect real-time use cases to increase or increase significantly over the next three years.

Real-time/low-latency use cases

Approximately what percentage of models deployed run in real-time or low-latency use cases/applications?



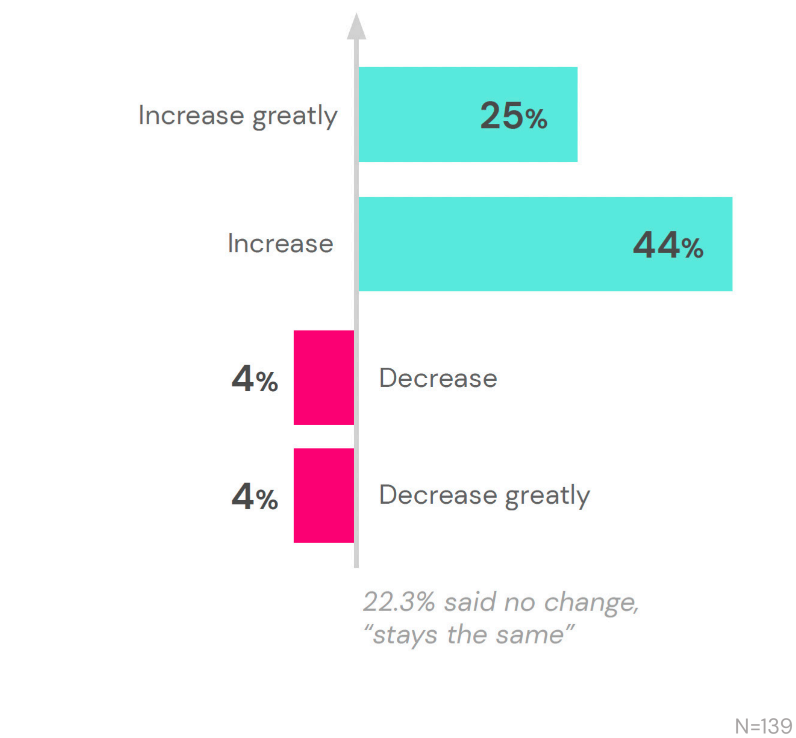
69%

Expect real-time use cases to increase or increase significantly over the next three years

The implications of this move to real time are significant for companies thinking about the infrastructure they will need to deploy to support Operational AI with real-time/low-latency uses at scale. Companies that are doing real-time AI in a massively scalable manner have adapted their processes and technology to separate what happens in batch and what happens in real time.

One very basic example: The technology stack centered on OLAP-based data warehouses that companies have used to support batch-oriented systems is very different from the stack required to support NLP in a smart home device or image recognition in an app that analyzes damage to an automobile for an insurance claim.

Real-time Use Cases Set to Increase



Verta Insights

Companies should plan their tech stack with the goal of supporting the use cases that will drive their business success. If those use cases include real time, ensure that your technology supports real-time Operational AI. And if you are already running batch use cases, ensure that you reevaluate your existing technology to understand its limitations around real time.

> "Real-time AI is essential for enabling intelligent applications."

Conrado Miranda
CTO, Verta

Operationalizing Models is Difficult

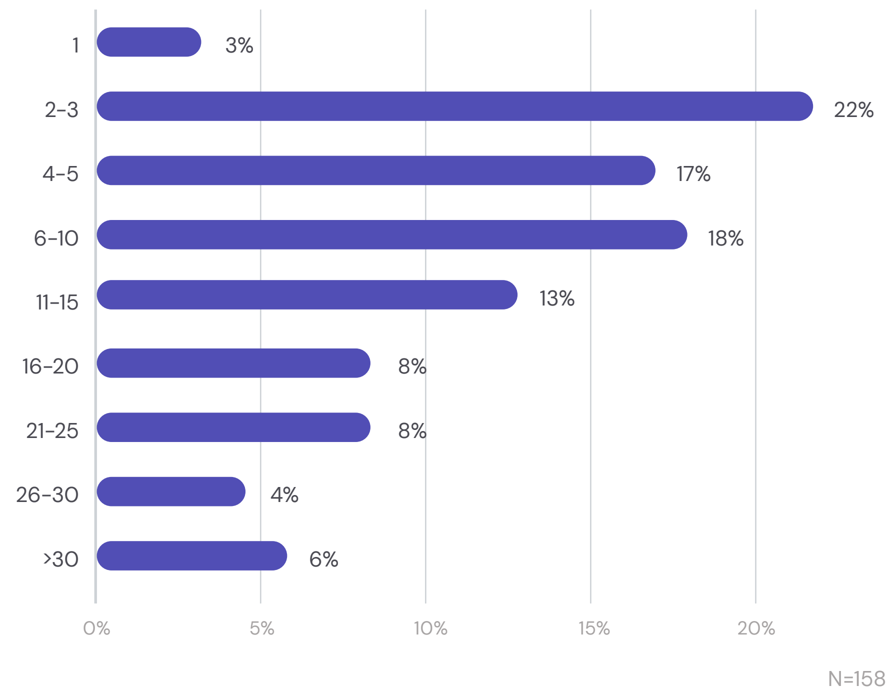
The study included a set of questions to gauge companies' performance in getting models into production. According to study participants:

- The average number of weeks to get a model into production is 12 weeks.
- The average number of AI projects put into production in the past quarter is 3.
- Nearly half of participants reported that their organizations didn't put any models into production in the past quarter, or they didn't know how many had been put into production.

Of course, models vary in their complexity and size, and different organizations use different kinds of models, so there is some level of natural variation based on the use case for a particular model and how much experimentation is required. Nevertheless, these statistics point to the "artisanal" nature of data science, and the fact that practitioners — who comprised the largest number of participants in the study — are focused on their work, building and training models, and not focused to the same extent on what happens to a model or a project once they hand it off.

Getting Models into Production

How long does it typically take to get a model into production? (# of weeks)



12 weeks

The average number of weeks to get a model into production

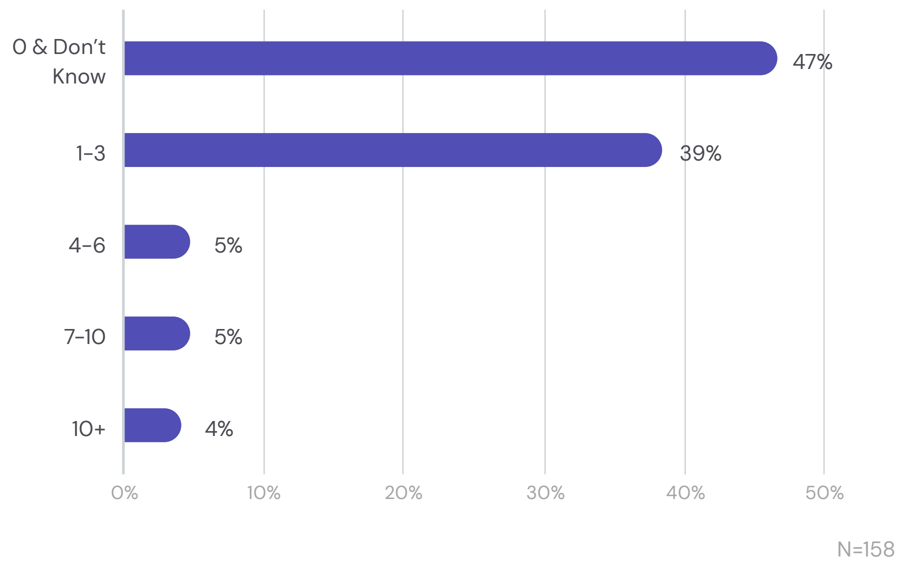
03

The average number of AI projects put into production in the past quarter

Notably, the level of effort and time required to put models into production are at odds with (or, more precisely, on a collision course with) trends pointing to the proliferation of models in general, and of real-time ML in particular. As the number of models that a business requires grows, unless the organization is able to accelerate the operationalization of models, inevitably the backlog will accumulate to a degree that becomes unsustainable. The acceleration in the adoption of real-time use cases also will put pressure on companies to significantly reduce their time to deliver models into production. The comparison here to software delivery times is instructive, where best-in-class companies are pushing out updates multiple times a day.

Projects Completed

How many successful AI projects, products or features did you ship last quarter?



"The lack of visibility among individuals and groups involved in building and training models, operationalizing models, and managing portfolios of models points to disconnects that companies need to bridge in order to achieve ML operational excellence."

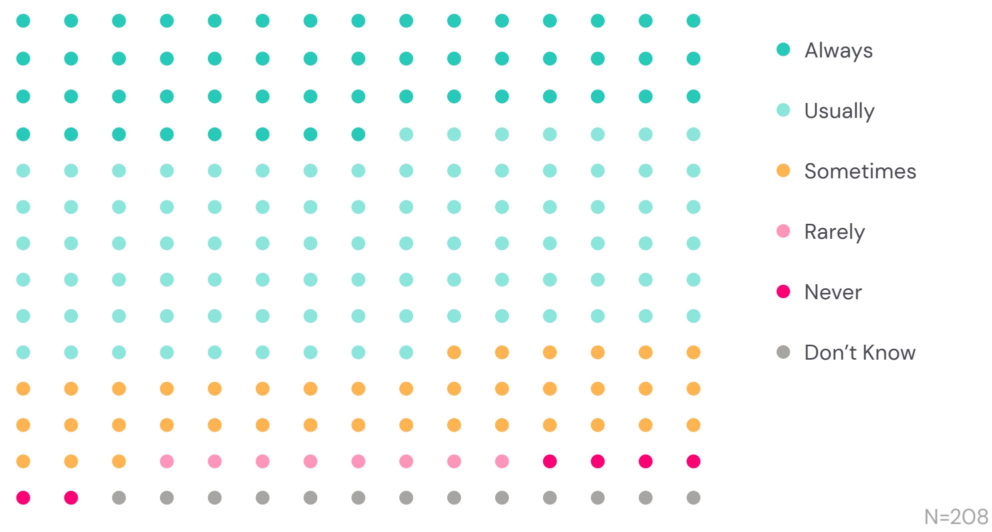
Meeta Dash
Vice President of Product, Verta

Comparing Leaders vs. Laggards

As part of the State of Machine Learning Operations Study, participants were asked to report on how frequently their organizations meet their financial targets.

The responses to this question allow us to establish a performance framework as a methodology for comparing the performance, characteristics and practices of "**leaders**" — who usually or always meet their targets — versus "**laggards**" — who sometimes, rarely or never do so.

How Often Company Meets Financial Goals



Performance

Comparing key performance metrics, the study showed that leaders consistently outperformed laggards by a significant margin with regard to three metrics. Specifically:

- Leaders are more than 2x likely (60% vs 25% for laggards) to successfully ship AI projects, products or features.
- Leaders are 2x more likely (49% vs 23%) to successfully launch AI/ML products or features into production.
- Leaders are 3x more likely (60% vs 20%) to meet service level agreements for uptime performance in operations.

It's notable that while leaders are outperforming laggards in key areas, nevertheless:

- Just one-third of leaders report that 70% or more of their organizations' projects make it into production; and,
- Only 21% - or 1 in 5 - of leaders have 80% of their models operating reliably in production.

These metrics are similar to statistics reported from other sources* that show less than 20% of models successfully operationalized and running in production. Comparing this level of performance against software development and reliability requirements of 99.999% uptime ("five 9s"), clearly the ML market, despite advances, remains a long way off from the high level of operational reliability and availability expected of critical business applications.

*E.g., Models Are Rarely Deployed: An Industry-wide Failure in Machine Learning Leadership, KDnuggets, January 17, 2022.

Performance Indicator

	LEADERS	LAGGARDS
Usually or always meet targets to ship AI projects, products or features (n=151)	60%	25%
Usually or always successfully launch AI/ML products/features into production (n=133)	49%	23%
Usually or always successfully meet SLA/Uptime/ Performance targets with models in production (n=133)	60%	20%
70% or more of the organizations' projects make it into production (n=151)	33%	23%
80%+ of organization's models successfully operate within production (n=133)	21%	18%

"

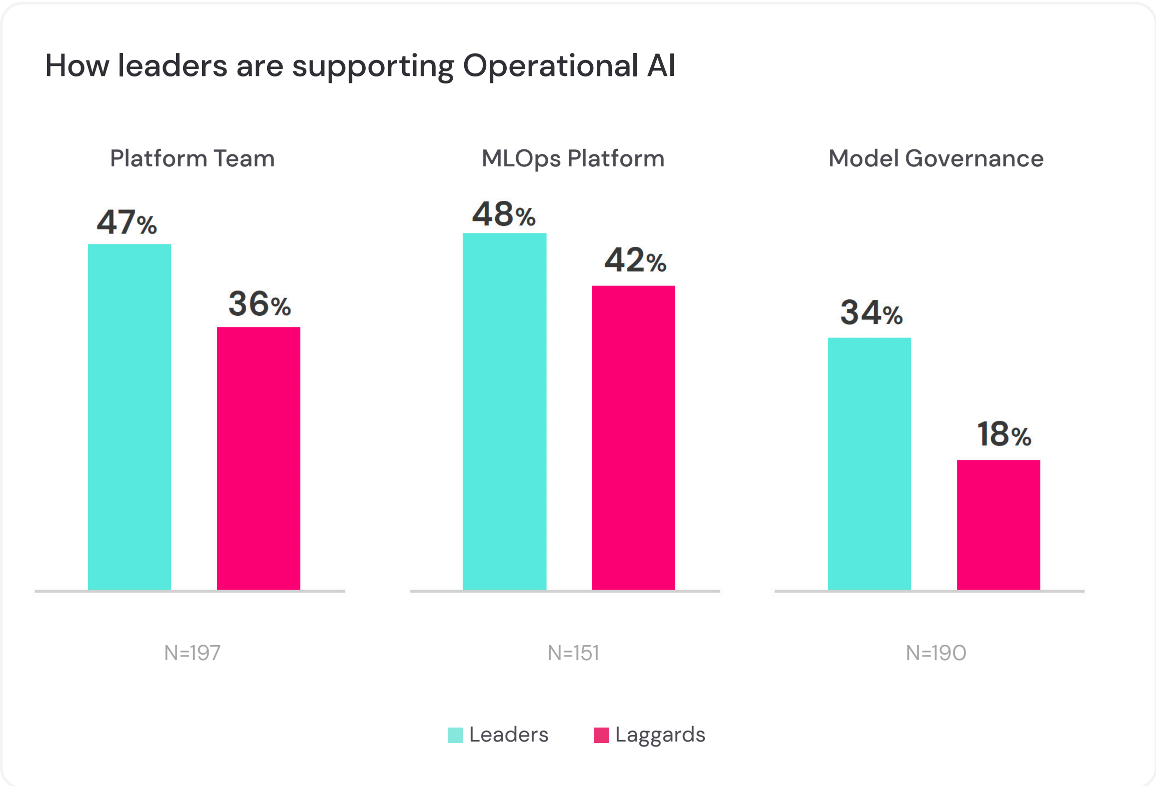
All industries are being rewritten with AI in their core DNA, and we have seen massive investments in data science. But leaders are clearly focusing now on ensuring they are able to operationalize their models, because that's where the value is generated and where companies realize the business impact of those investments."

Manasi Vartak
CEO and Founder, Verta

Characteristics

The study explored characteristics that distinguish leaders from laggards, and several trends emerged. Leaders are more likely to have put in place an ML platform team and an MLOps platform, and also are more likely to have a formal structure around model governance and oversight.

Companies use different nomenclature for their ML platform teams (see sidebar "An ML Platform Team by Any Other Name"), but their purview covers identifying, integrating, and maintaining the set of tools that the data science team uses (similar to how an IT team would support software developers).



"An ML Platform Team by Any Other Name"

Study participants reported a wide variety of different names for their organization's ML Platform Team:

- ADA team
- AI Application
- AI team
- AI/ML Engineering
- AIOps Team
- CDO AI/ML Platform Team
- Data & Analytics
- Data & ML
- Data Engineering
- Data Insights
- Data Office
- Data Science Team
- Data Support Team
- DevOps
- Emerging Tech Support
- IT Platform Engineering
- ITS support team
- Machine Learning
- Machine Learning Ops
- ML Engineering
- ML Platform Team
- ML Production team
- Ml Research Group
- MLOps
- Model Deployment Team
- Model Management
- Support team
- Systems Engineering

The platform team can handle the "heavy lifting" associated with managing the infrastructure so that Data Science can focus on their core competencies around model building and training. Platform teams also often are the first line of defense in the event of a production incident for real-time models.

Typically these are small teams, numbering up to ten specialists, or smaller for an organization just getting started with ML. However, industry experience has shown that these teams have an outsized impact on the success and scalability of AI and data science initiatives.

In terms of model governance, this responsibility can fall on different roles. Financial organizations may assign model oversight to an enterprise risk team. A technology company may have an ethics council. However, as the present study shows, more companies are tasking a specific group with the role of ensuring that the models that run the business have governance and responsible AI practices in place. (See sidebar: "Who Is Doing Model Governance?")

> **The number one differentiator that we're seeing at companies that are using AI successfully throughout their organizations is that they all have ML platform teams."**

Conrado Miranda
CTO, Verta

"Who Is Doing Model Governance?"

Study participants that reported having a model governance/ethics council at their organization were asked what that team is called. Responses included:

- ADA team
- AI Ethics Council
- AI Governance
- Chief Data & Analytics office
- Cloud Intake Governance Team
- Data Ethics and Governance
- Data Governance
- Data Science Team
- DevRel
- Ethics
- Ethics Committee
- Global Data Governance
- IA Team
- Model Governance
- Model GRC
- Model Risk Governance Committee (MRGC)
- Privacy team
- Responsible AI council

Retraining Practices

Finally, the study looked at differences in how leaders and laggards approach model retraining and monitoring.

Starting with retraining, we asked about the frequency of retraining and their approach and process for retraining. Regarding Retraining Approach and Process, it was notable that leaders were more advanced in terms of automation, by a fairly significant margin, and also marginally more mature in terms of process. In terms of Retraining Practices, fewer leaders reported retraining on a weekly basis than laggards, while leaders were more likely to retrain on a monthly basis.

Retraining Frequency

	LEADERS	LAGGARDS
Daily	1%	2%
Weekly	19%	25%
Monthly	41%	30%
Quarterly	17%	28%
Annually	8%	6%
Don't know	14%	9%

N=196

Approach to Retraining

	LEADERS	LAGGARDS
Fully automated	8%	2%
Highly automated	31%	21%
Partially automated	42%	45%
Manual	15%	23%
No automated	5%	9%

N=196

Retraining Process

	LEADERS	LAGGARDS
Systematized tools	12%	11%
Metrics/controls	22%	19%
Defined/ documented	20%	28%
Somewhat repeatable	36%	23%
Ad hoc	10%	19%

N=196

Monitoring Practices

Looking at Monitoring Approach, leaders appear to be somewhat more advanced in terms of automation. But with regard to Monitoring Process, leaders were significantly more advanced than laggards.

These results are suggestive of a correlation: Organizations that have more mature practices and processes around retraining and monitoring tend to retrain less frequently. One possible explanation for this could be that companies that have more diagnostics around monitoring retrain on an as-needed basis when they detect an issue with a model. As opposed to a company that has less diagnostics around monitoring and therefore less visibility into their models' performance, and therefore has adopted a scheduled approach to retraining in hopes of updating their model before issues occur.

Verta Insights

The results of this section of the survey highlight again that ML is still very much in the early stages of maturation, with even leaders reporting relatively low levels of full automation and implementation of systematized tools.

Approach to Monitoring

	LEADERS	LAGGARDS
Fully automated	10%	9%
Highly automated	24%	17%
Partially automated	45%	47%
Manual	12%	15%
No automation	10%	11%

N=196

Monitoring Process

	LEADERS	LAGGARDS
Systematized tools	10%	2%
Metrics/controls	25%	17%
Defined/ documented	28%	40%
Somewhat repeatable	24%	28%
Ad hoc	13%	13%

N=196

It's About Operational AI Excellence Now

AI is creating new opportunities for growth, competitive advantage, and customer experience. But the gains in operational performance appear to be more of an imperative for keeping pace than they are for getting ahead.

To elaborate, AI is known to be a differentiator in the market. Organizations are introducing new revenue streams by delivering AI-enabled products and services. But while these intelligent systems, applications and equipment are driving incremental revenue gains, they are becoming more of a cost of doing business rather than an investment in competitive advantage.

Looking at how companies currently are prioritizing their ML investments, it's clear that leaders believe that the next level of competitive advantage is going to come from being able to effectively operationalize models and manage models in production. At this stage, companies are less worried about getting ahead by introducing cool new intelligent applications, and more worried about being left behind by failing to achieve Operational AI excellence.

"As AI adoption and real-time use cases scale dramatically, organizations will need to augment their technology stack to include operational AI infrastructure if they intend to achieve top-line benefits through intelligent equipment, systems, products and services."

Manasi Vartak
CEO and Founder, Verta

Demographics

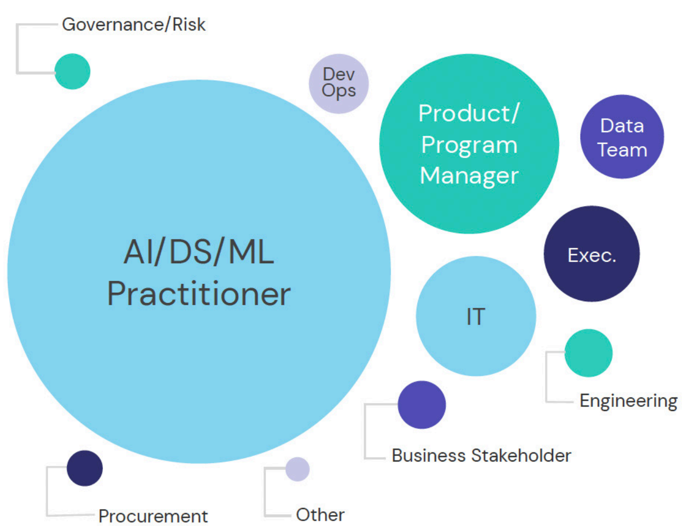
Job Role and Responsibilities

This study was based on a survey conducted online in July-August 2022. Total respondents included 1,000+ technology practitioners, but for the purposes of the study we included only the results from 208 validated MLOps and adjacent stakeholders located in the US and Canada.

The study participant group included broad representation across different roles and responsibilities in the enterprise, and the participants were generally weighted toward practitioners.

Job Role

N=208



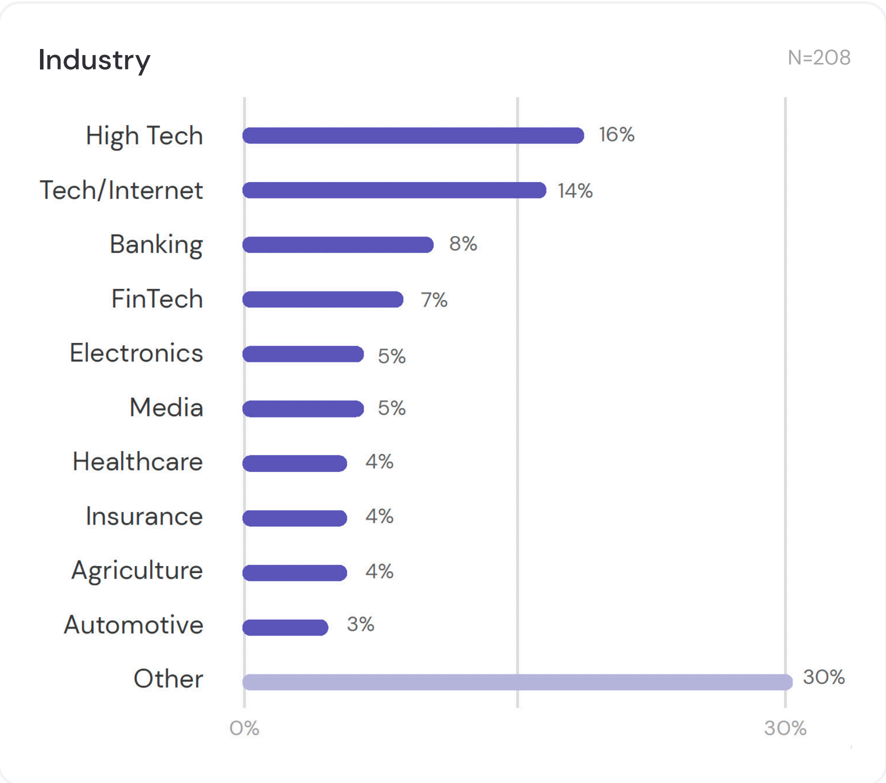
Job Responsibilities

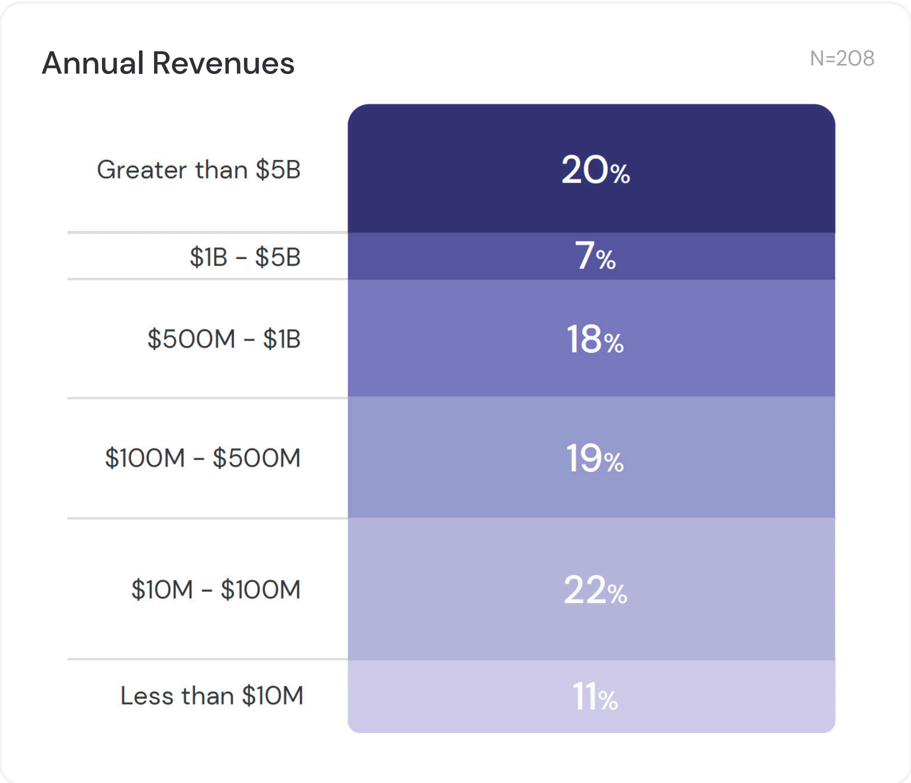
N=208

Manager-level	28%
Technical Fellow | Chief	20%
Individual Contributor	18%
Director-level	11%
Consultant	6%
VP-level	5%
President | CEO | Chairman	4%
C-Level (CDO, CIO, CTO, CFO)	3%
Owner | Founder	2%
Other	3%

Industry and Company Size

The greatest number of participants came from companies in the high tech or internet sectors, but respondents also came from companies across a range of different sectors that are implementing AI, including digital-native companies that have based their entire business around AI.





High Tech encompasses semiconductors and electronic components. Tech/Internet includes digital-native companies. Electronics includes consumer electronics devices.

Appendix

What is different about Operational AI?

Operational AI, by definition, is focused on the operationalization of models and the management of models in production. Operational AI is distinguished from Experimentation, which is focused on building and training models.

Operational AI encompasses a few characteristics

1. **Putting models to use:** First, it means putting models to use: taking models that have been in experimentation mode and actually converting them into products or services that the broader business can use.
2. **Real time:** We also use Operational AI to indicate some level of real-time inference versus batch operations (although Operational AI doesn't necessarily preclude batch).
3. **Continuous operations:** Also, Operational AI is not just about the first time that you put models into products or services. It's about keeping them continually operating to a high standard.

Reference Architecture for Operational AI

In terms of how companies can think through their options for enabling ML operations, it can be useful to put structure around the groupings of different technologies by using a reference architecture.

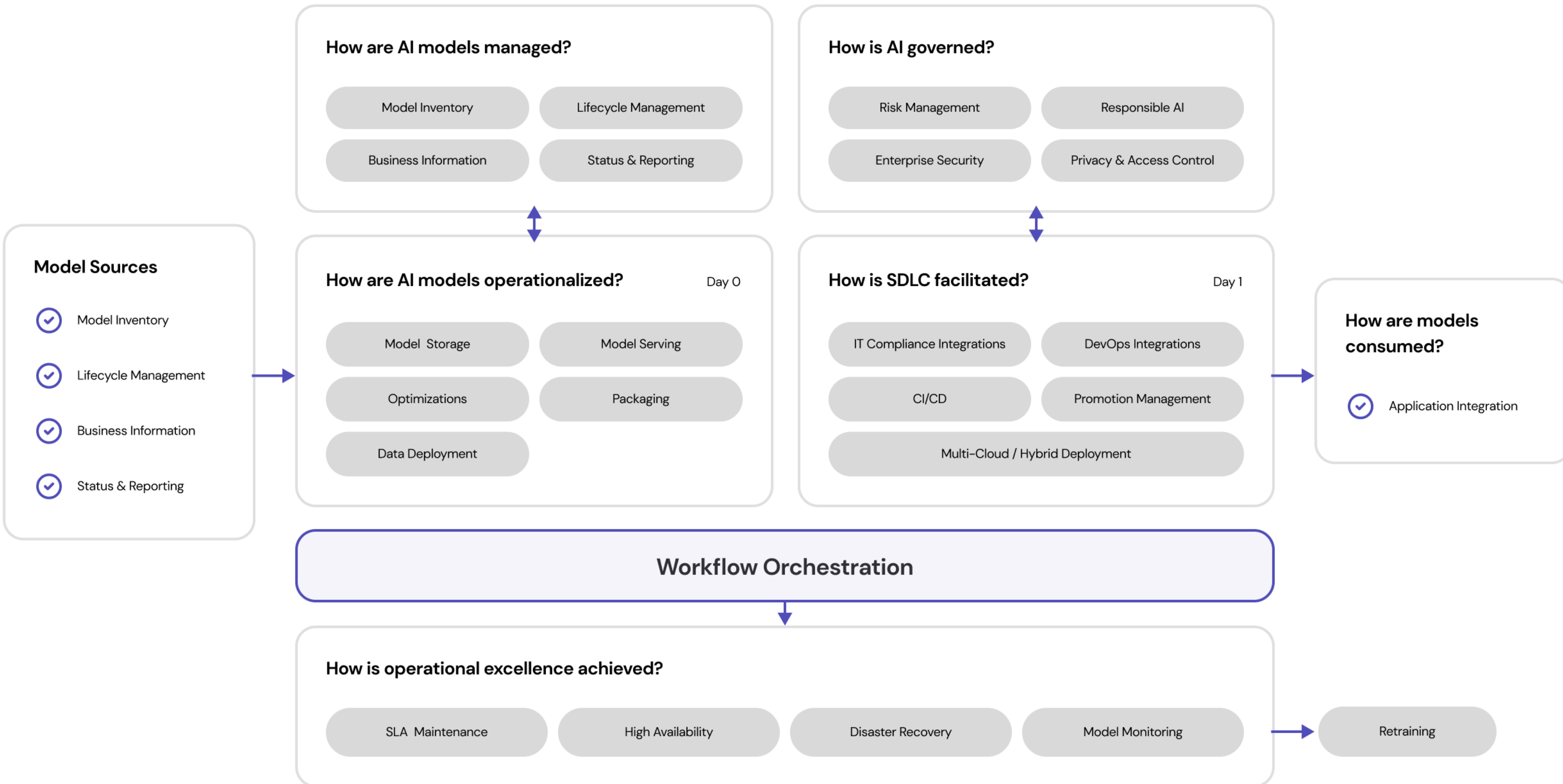
For Experimentation, Gartner has developed their Reference Architecture to provide a roadmap for companies starting their journey on AI and ML. The first step for these companies is to put in place tools to build and train models, and the kinds of questions that Gartner posits these companies should be asking themselves are around how their data is managed, consumed, and governed. The reference architecture then shows the individual building blocks that you need in place to cover an MVP reference architecture for experimentation, whether those are tools that you build internally or that you deploy from an external vendor.

With regard to Operational AI, this is a less mature market than Build and Train. In fact, Gartner has stated that the Operational AI market is 1-5% penetrated. To provide a roadmap for enabling model operations, Verta has worked with practitioners and others in the space to develop a reference architecture for Operational AI. As with Gartner's model, the Reference Architecture for Operational AI highlights the questions that companies should be asking themselves as they plan their strategy for enabling model operations.

> "The Reference Architecture for Operational AI is a blueprint that helps enterprise architects and leaders of data science and ML teams understand the requirements and challenges of enabling model operations, both in terms of the tooling and the team."

Manasi Vartak
CEO and Founder, Verta

Reference Architecture for Operational AI



What do we mean by "real time"?

"Real time" implies

1. **Fast response:** Real time means you're going to respond very quickly to a request, usually because a customer depends on getting a response "instantly." You can do that via an API or via a streaming system like Kafka.
2. **On-device:** Real time also can mean that you're "on-device" and you're doing things as you receive the data.

Real-time differs from batch/offline inference in several important ways

1. **Amount of data:** One difference is the amount of data you are typically processing:
    - Batch usually means you are processing gigabytes of data, millions of records, and they can be arbitrarily complex. Because you don't need an answer instantly, these are going to run over many minutes or hours, sometimes over many days.
    - In real time, the individual request that you're operating on is not going to have a large amount of data – perhaps a few megabytes at most. You might be serving a lot of customers at the same time, but the data involved in that individual request is going to be relatively small.
2. **Response time:** A big difference is the user's expectations for response time – an issue that drives a lot of system design:
    - If we don't need to respond to the customer in less than an hour, we would use a very different system design than if we need to respond to the customer in a few milliseconds.
    - The required response time will impact the way that you architect your databases, how you instrument them to get monitoring, for instance. APM matters for real time because there are more places for it to break down because you're doing millions of smaller requests as opposed to doing one single but massive request.
3. **System design/infrastructure:** The infrastructure for batch needs to be very different from the infrastructure for real time, because you build an infrastructure to serve a particular kind of workload.
    - If you have three days, then you don't need super efficient code. You can just read your records one at a time.
    - For real time, you need very different data structures. For instance, there are online analytical processing (OLAP) systems that are offline systems for analytics, and then there are online transaction processing (OLTP) systems, for transaction processing (i.e., for real time). The system that lets you get money at the ATM is OLTP, because if you were to use an OLAP-based analytics system in an ATM, you're not getting your money anytime soon.
4. **Where your data lives:** Operational data, online data and offline data live in different places:
    - Your operational data might live in Aurora, which means it's the data that you want to efficiently access.
    - Your online data might live in a Cassandra - it's in-memory, where you don't need to go to a disk and read the data. If you want millisecond responses, you're going to have data sitting in memory so you can respond very quickly.
    - Your offline data is going to be sitting on disk somewhere where it can take its time reading and processing.

Verta

About Verta

Verta provides AI/ML model management and operations software that helps enterprise data science teams to manage inherently complex model-based products.

Verta's production-ready systems help data science and IT operations teams to focus on their strengths and rapidly bring AI/ML advances to market. Based in Palo Alto, Verta is backed by Intel Capital and General Catalyst.

For more information visit

www.verta.ai

@VertaAI

linkedin.com/company/vertaai

Credits

The State of Machine Learning Operations 2022 research project was conducted by Verta Insights, the research group at Verta, Inc.

VERTAINSIGHTS RESEARCH

About Verta Insights

Verta Insights is the research group at Verta, a leading provider of Artificial Intelligence (AI) model management and operations solutions. Verta Insights conducts research into trends in the AI and machine learning space, and delivers insights to assist AI/ML practitioners and executive leaders to prepare their organizations for the AI-enabled intelligent future.