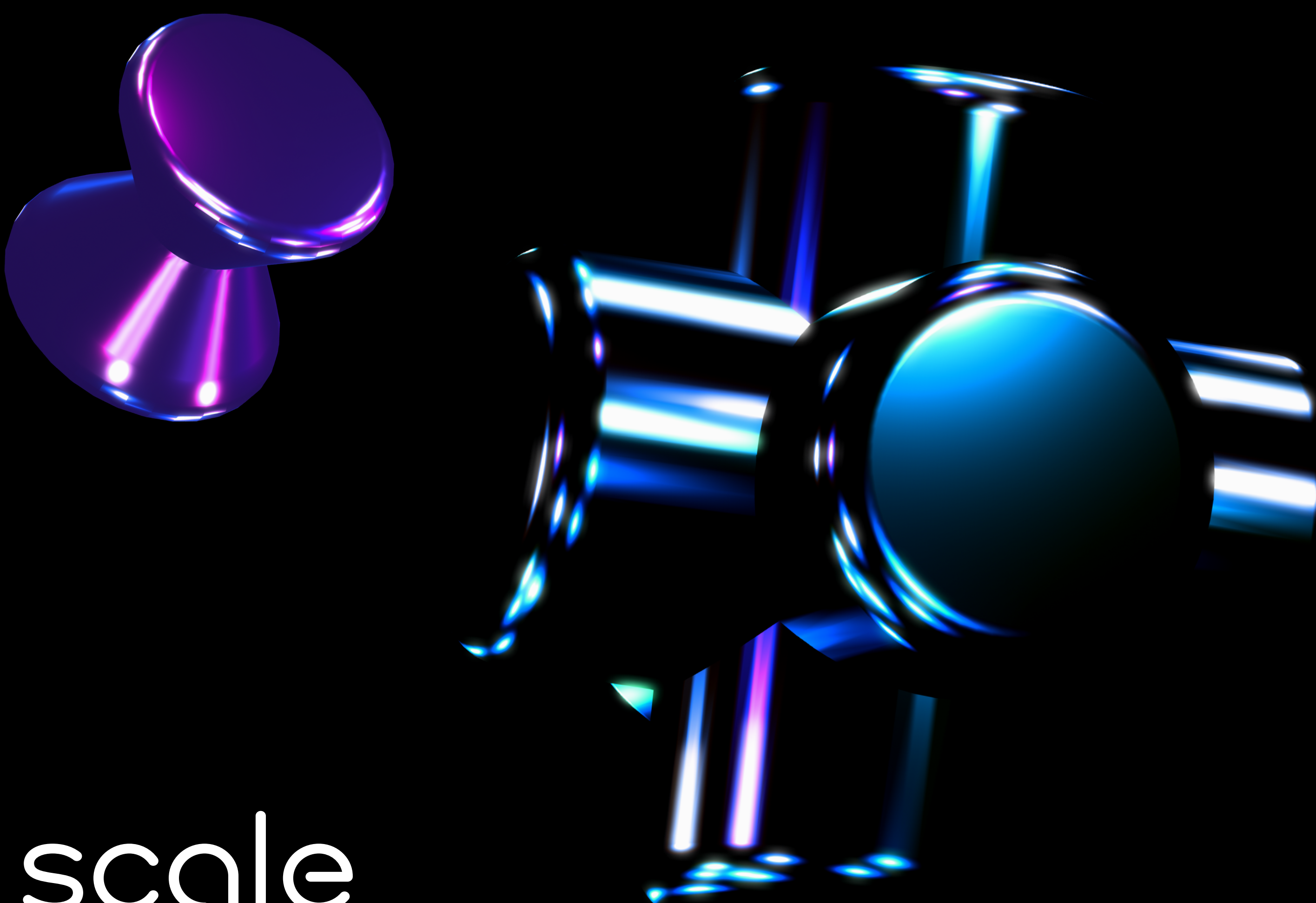
zeitgeist

AI Readiness Report



scale

Table of Contents

Introduction

1st edition of the annual AI Readiness Report

At Scale, our mission is to accelerate the development of AI applications to power the most ambitious AI projects in the world. That's why today we're introducing the Scale Zeitgeist: AI Readiness Report, a survey of more than 1,300 ML practitioners to uncover what's working, what's not, and the best practices for ML teams and organizations to deploy AI for real business impact.

Today, every industry – from finance, government, healthcare, and everything in between – has begun to understand the transformative potential of AI and invest in it across their organizations. However, the next question is: are those investments succeeding? And are the organizations set up in the right way to foster meaningful outcomes? Is your business truly "AI-ready?"

We designed the AI Readiness Report to explore every stage of the ML lifecycle, from data and annotation to model development, deployment, and monitoring, in order to understand where AI innovation is being bottlenecked, where breakdowns occur, and what approaches are helping companies find success. Our goal is to continue to shed light on the realities of what it takes to unlock the full potential of AI for every business. We hope these insights will help empower organizations and ML practitioners to clear their current hurdles, learn and implement best practices, and ultimately use AI as a strategic advantage.

"In the world of AI, there are a lot of people focused on trying to solve problems that are decades away from becoming a reality, and not enough people focused on how we can use this technology to solve the world's problems today. It can feel like a leap of faith to invest in the necessary talent, data, and infrastructure to implement AI, but with the right understanding of how to set ML teams up for success, companies can start reaping the benefits of AI today. Thanks to all the ML experts who shared their insights for this report, we can help more companies harness the power of AI for decades to come."

Alexandr Wang — Founder & CEO, Scale

lenges Key Challenges Key Ta
ges Key Challenges Key Challe

Key Challenges

Data

For all teams working on ML projects, data quality remains a challenge.

Factors contributing to data quality issues include data complexity, volume, and scarcity.

Furthermore, annotation quality and a team's ability to effectively curate data for its models affect speed of deployment.

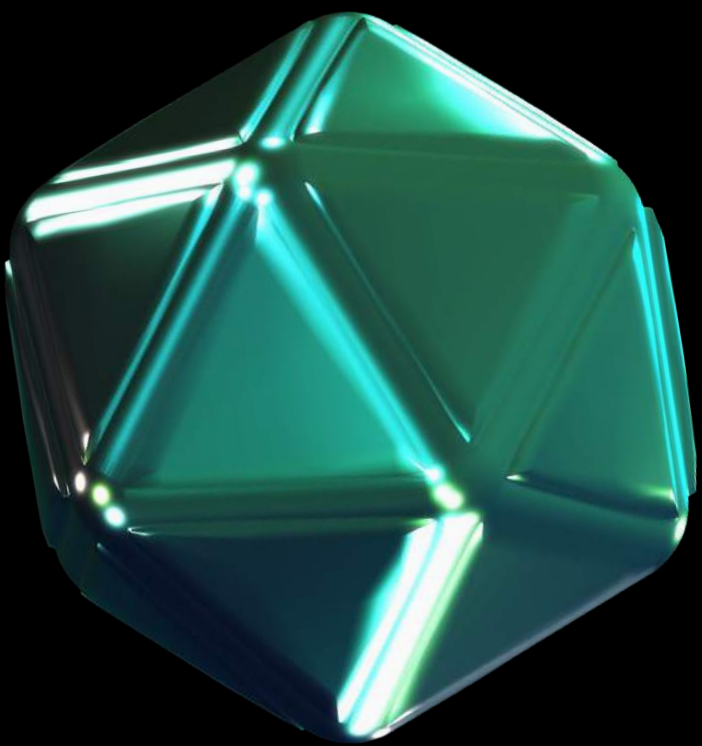
Model Development, Deployment, and Monitoring

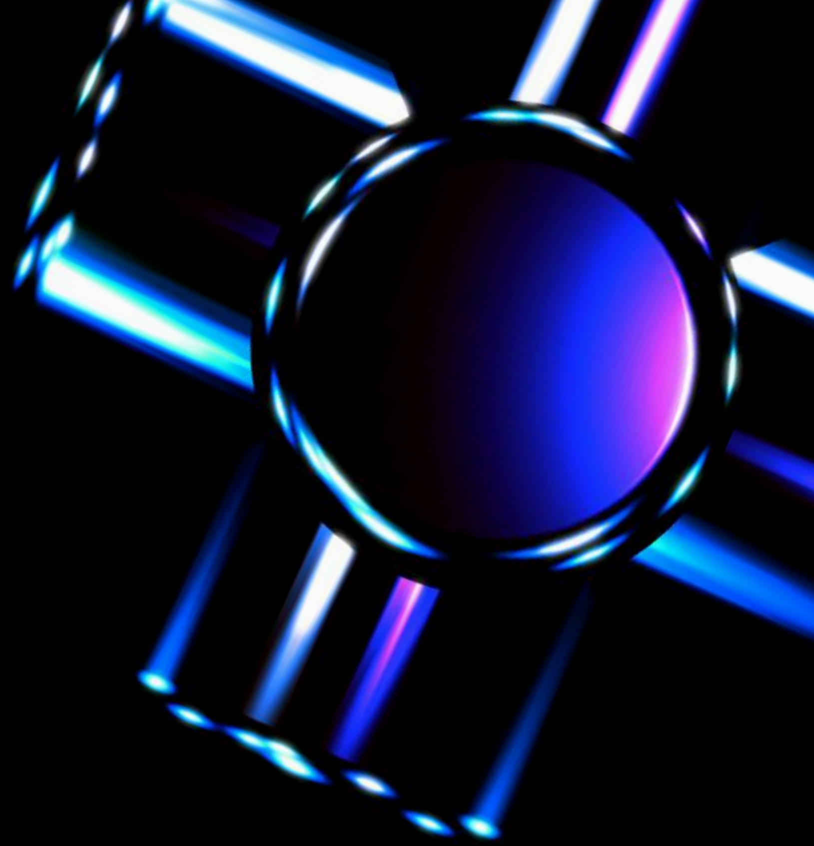
Feature engineering is a top challenge for all teams working in ML.

Smaller companies are least likely to evaluate models by their business impact

The majority of ML practitioners cite deployment at scale as a challenge.

Identifying issues in models takes longer for larger companies.

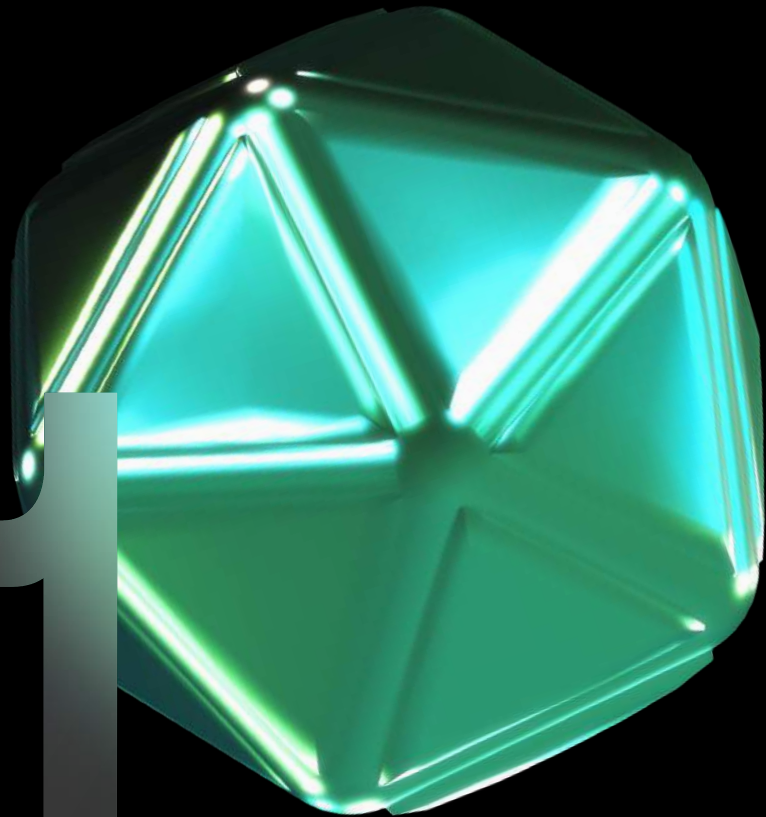




Data

Data embodies the foundation for all machine learning. Many teams spend most of their time on this step of the ML lifecycle, collecting, analyzing, and annotating data to ensure model performance. In this chapter, we explore some of the key AI challenges ML teams encounter when it comes to data for their initiatives and discuss some of the latest AI trends, including best practices and guidance from leading ML teams for data.

01

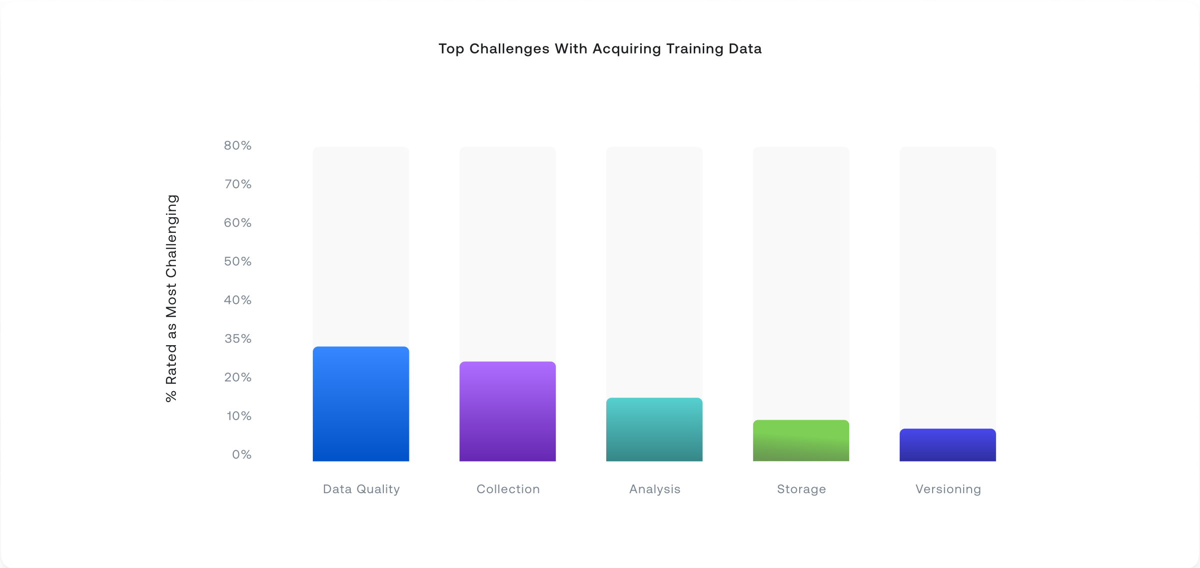


Data Challenges

Data quality is the most challenging part of acquiring training data.

When it comes to acquiring training data, the barriers to success for ML teams include challenges with collection, quality, analysis, versioning, and storage. Respondents cited data quality as the most difficult part of acquiring data (33%), closely followed by data collection (31%). These problems have a significant downstream impact on ML efforts because teams often cannot model effectively without quality data.

Challenge 1



• **FIGURE 1**: MOST RESPONDENTS CITED DATA QUALITY AS THE MOST CHALLENGING ASPECT OF ACQUIRING TRAINING DATA, FOLLOWED BY DATA COLLECTION.

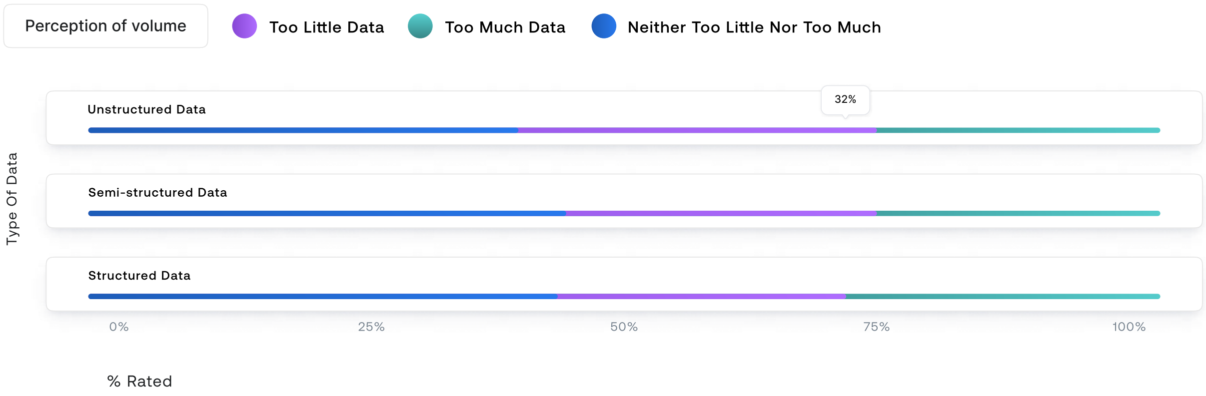
> We have to spend a lot of time sourcing—preparing high-quality data before we train the model. Just as a chef would spend a lot of time to source and prepare high-quality ingredients before they cook a meal, a lot of the emphasis of the work in AI should shift to systematic data preparation.

Andrew Ng
Founder and CEO, Landing AI

Factors contributing to data quality challenges include variety, volume, and noise.

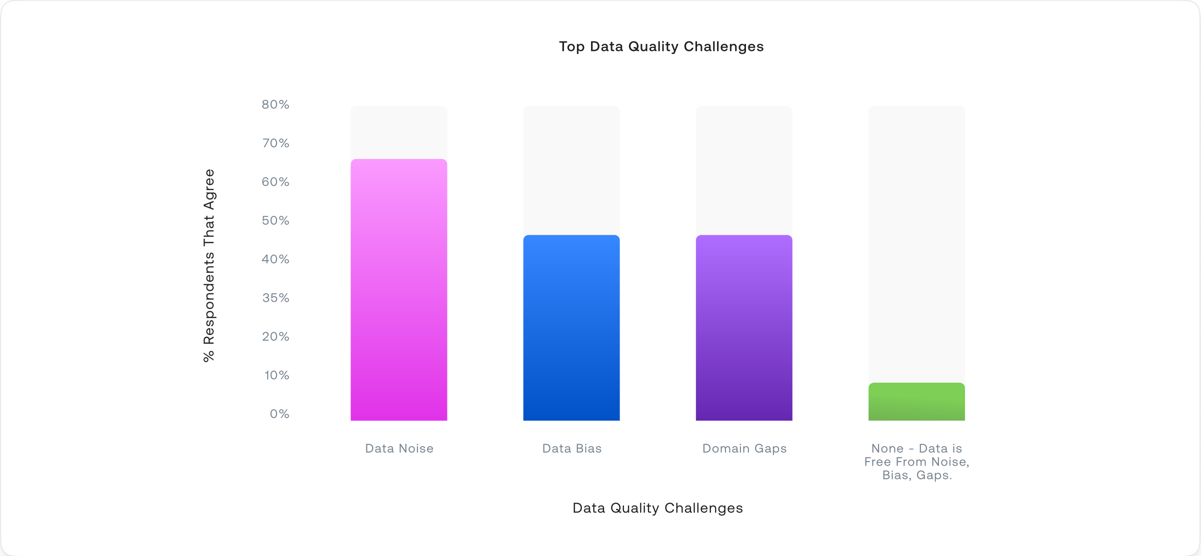
To learn more about what factors contribute to data quality, we explored how data type affects volume and variety. Over one-third (37%) of all respondents said they do not have the variety of data they need to improve model performance. More specifically, respondents working with unstructured data have the biggest challenge getting the variety of data they need to improve model performance. Since a large amount of data generated today is unstructured, it is imperative that teams working in ML develop strategies for managing data quality, particularly for unstructured data.

Challenge 2



• **FIGURE 2**: RESPONDENTS WORKING WITH UNSTRUCTURED DATA ARE MORE LIKELY THAN THOSE WORKING WITH SEMI-STRUCTURED OR STRUCTURED DATA TO HAVE TOO LITTLE DATA.

The majority of respondents said they have problems with their training data. Most (67%) reported that the biggest issue is data noise, and that was followed by data bias (47%) and domain gaps (47%). Only 9% indicated their data is free from noise, bias, and gaps.



• **FIGURE 3**: THE MAJORITY OF RESPONDENTS HAVE PROBLEMS WITH THEIR TRAINING DATA. THE TOP THREE ISSUES ARE DATA NOISE (67%), DATA BIAS (47%), AND DOMAIN GAPS (47%). [NOTE: SUM TOTAL DOES NOT ADD UP TO 100% AS RESPONDENTS WERE ASKED TO STACK-RANK OPTIONS.]

Five tips for data-centric AI development: Make labels consistent, use consensus labeling to spot inconsistencies, clarify labeling instructions, toss out noisy examples (because more data is not always better), and use error analysis to focus on a subset of data to improve.

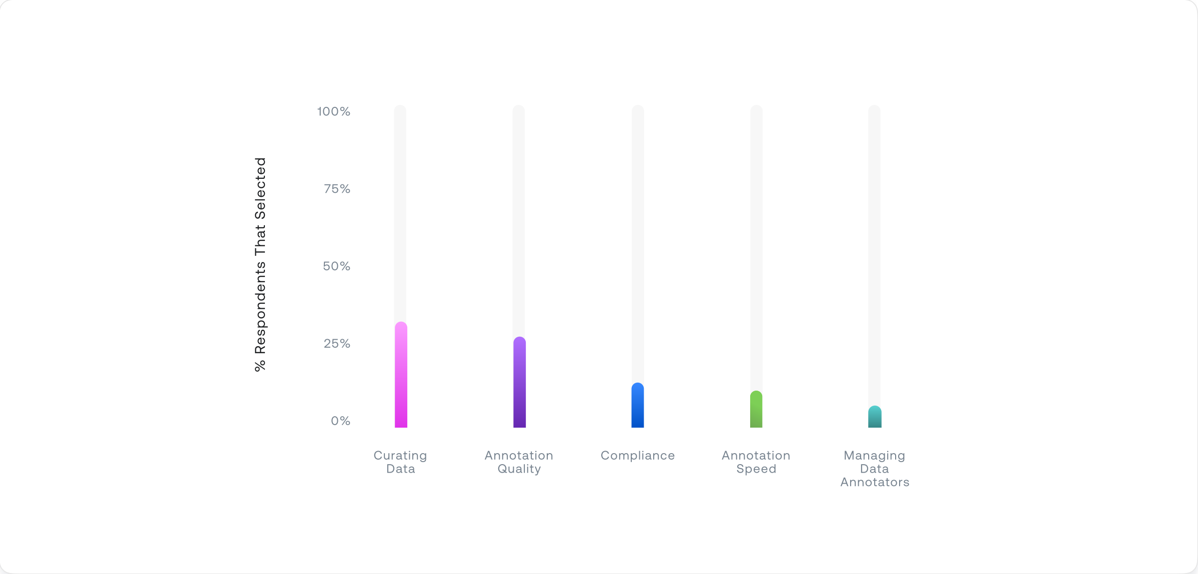
Andrew Ng
Founder and CEO, Landing AI

Curating data to be annotated and annotation quality are the top two challenges for companies preparing data for training models.

When it comes to preparing data to train models, respondents cited curating data (33%) and annotation quality (30%) as their top two challenges. Curating data involves, among other things, removing corrupted data, tagging data with metadata, and identifying what data actually matters for models. Failure to properly curate data before annotation can result in teams spending time and budget annotating data that is irrelevant or unusable for their models. Annotating data involves adding context to raw data to enable ML models to generate predictions based on what they learn from the data. Failure to annotate data at high quality often leads to poor model performance, making annotation quality of paramount importance.

Challenge 3

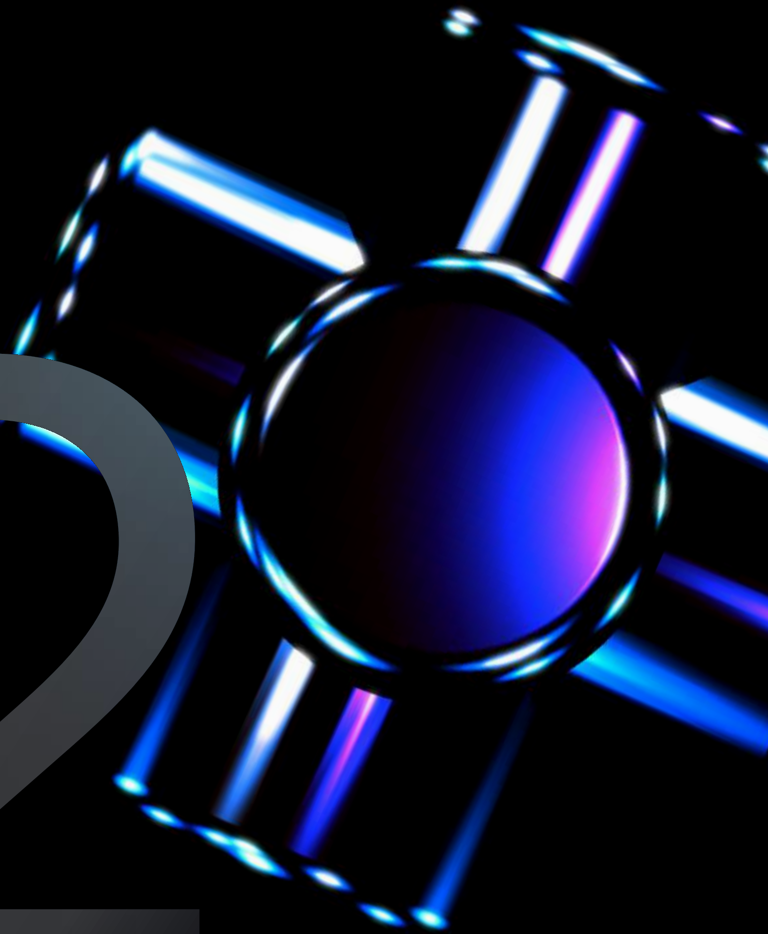
Top Challenges For Preparing Data For Model Training



• **FIGURE 4**: CURATING DATA AND DATA QUALITY ARE THE TOP CHALLENGES FOR COMPANIES PREPARING DATA FOR TRAINING MODELS. [NOTE: SUM TOTAL DOES NOT ADD UP TO 100% AS RESPONDENTS WERE ASKED TO STACK-RANK OPTIONS.]

> It's a lot harder for us to apply automation to the data that we obtain, because we get it from external service providers in the healthcare industry who don't necessarily have maturity on their side in terms of how they think about data feeds. So we have to do a lot of manual auditing to ensure that our data that's going into these machine learning models is actually the kind of data that we want to be using.

Vishnu Rachakonda,
ML Engineer, One Hot Labs



02

Data Best Practices

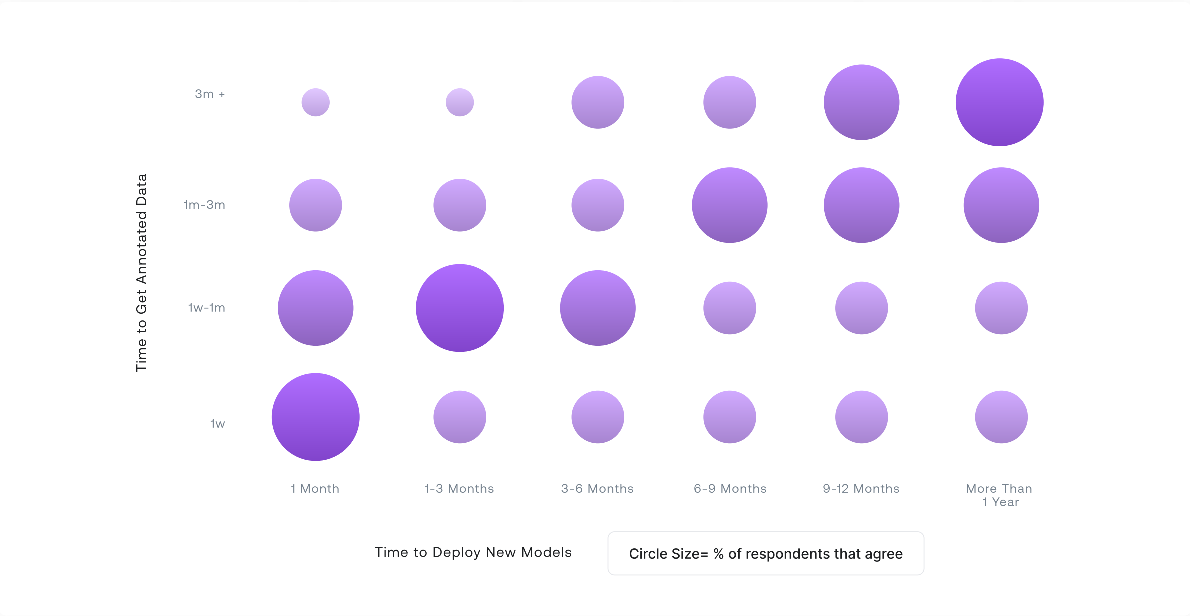
Part 2 - Data Best Practices

Companies that invest in data annotation infrastructure can deploy new models, retrain existing ones, and deploy them into production faster.

Our analyses reveal a key AI readiness trend: There's a linear relationship between how quickly teams can deploy new models, how frequently teams retrain existing models, and how long it takes to get annotated data. Teams that deploy new models quickly (especially in less than one month) tend to get their data annotated faster (in less than one week) than those that take longer to deploy. Teams that retrain their existing models more frequently (e.g., daily or weekly) tend to get their annotated data more quickly than teams that retrain monthly, quarterly, or yearly.

As discussed in the challenges section, however, it is not just the speed of getting annotated data that matters. ML teams must make sure the correct data is selected for annotation and that the annotations themselves are of high quality. Aligning all three factors—selecting the right data, annotating at high quality, and annotating quickly—takes a concerted effort and an investment in data infrastructure.

Best Practice 1



• **FIGURE 5**: TEAMS THAT GET ANNOTATED DATA FASTER TEND TO DEPLOY NEW MODELS TO PRODUCTION FASTER. [NOTE: SOME RESPONDENTS INCLUDED TIME TO GET ANNOTATED DATA AS PART OF THE TIME TO DEPLOY NEW MODELS.]

• **FIGURE 6**: TEAMS THAT GET ANNOTATED DATA FASTER TEND TO BE ABLE TO RETRAIN AND DEPLOY EXISTING MODELS TO PRODUCTION MORE FREQUENTLY. [NOTE: SIMILAR TO FIGURE 5, SOME RESPONDENTS INCLUDED TIME TO GET ANNOTATED DATA AS PART OF THE TIME TO RETRAIN. IN THE CASE OF RETRAINING, HOWEVER, TEAMS CAN STILL RETRAIN MONTHLY, WEEKLY, OR EVEN DAILY, BY ANNOTATING DATA IN LARGE BATCHES.]

> You identify some scenes that you want to label, which may involve some human-in-the-loop labeling, and you can track the turnaround time, given a certain amount of data. And for model training, you can try to optimize with distributed training to make it faster. So you have a pretty good understanding of how long it takes to train a model on the amount of data that you typically train a model on. And the remaining parts, like evaluation, should be fairly quick.

Jack Guo,

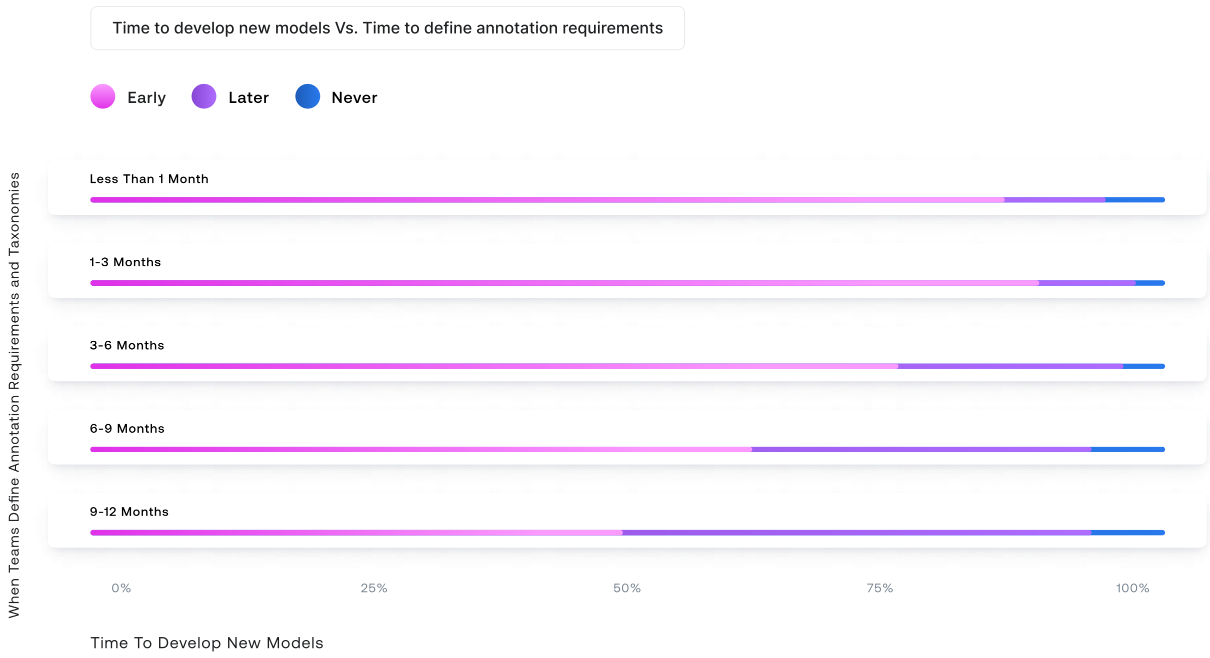
Head of Autonomy Platform, Nuro

ML teams that work closely with annotation partners are most efficient in getting annotated data.

The majority of respondents (81%) said their engineering and ML teams are somewhat or closely integrated with their annotation partners. ML teams that are not at all engaged with annotation partners are the most likely (15% v 9% for teams that work closely) to take greater than three months to get annotated data. Therefore, our results suggest that working closely is the most efficient approach.

Furthermore, ML teams that define annotation requirements and taxonomies early in the process are likely to deploy new models more quickly than those that are involved later. ML teams are often not incentivized or even organizationally structured to work closely with their annotation partners. However, our data suggests that working closely with those partners can help ML teams overcome challenges in data curation and annotation quality, accelerating model deployment.

Best Practice 2



• **FIGURE 7**: TEAMS THAT DEFINE ANNOTATION REQUIREMENTS AND TAXONOMIES EARLY IN THE PROCESS ARE LIKELY TO DEPLOY NEW MODELS MORE QUICKLY THAN THOSE THAT DO SO LATER IN THE PROCESS.

Document your labeling instructions clearly. Instructions illustrated with examples of the concept, such as showing examples of scratched pills, of borderline cases and near misses, and any other confusing examples—having that in your documentation often allows labelers to become much more consistent and systematic in how they label the data.

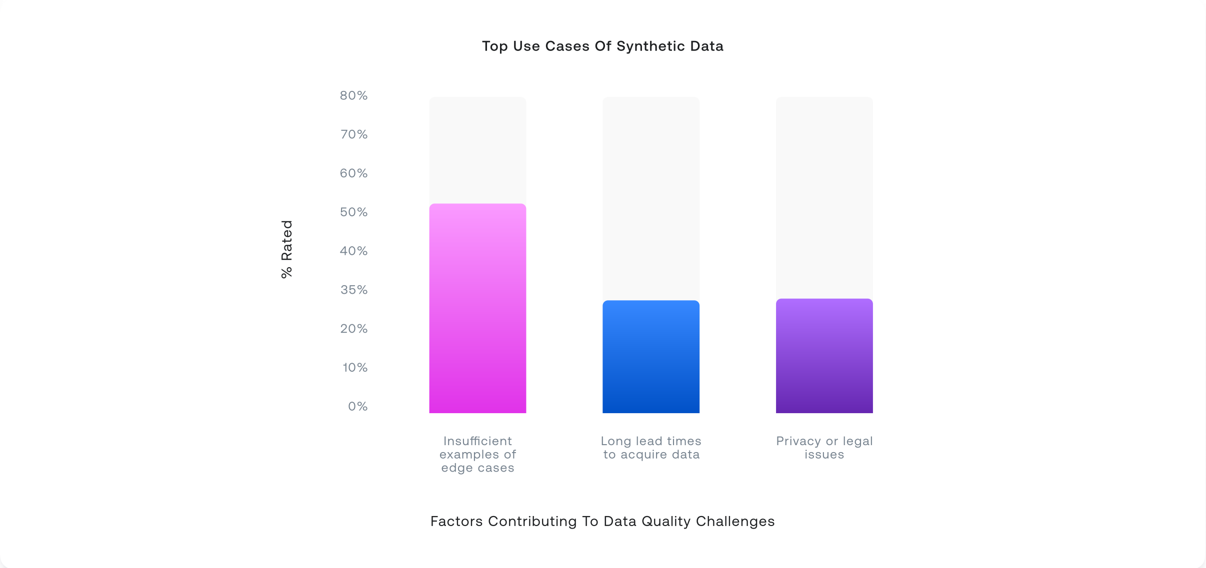
Andrew Ng,

Founder and CEO, Landing AI

To address data quality and volume challenges, many respondents use synthetic data.

Among the whole sample, 73% of respondents leverage synthetic data for their projects. Of those, 51% use synthetic data to address insufficient examples of edge cases from real-world data, 29% use synthetic data to address privacy or legal issues with real-world data, and 28% use synthetic data to address long lead times to acquire real-world data.

Best Practice 3

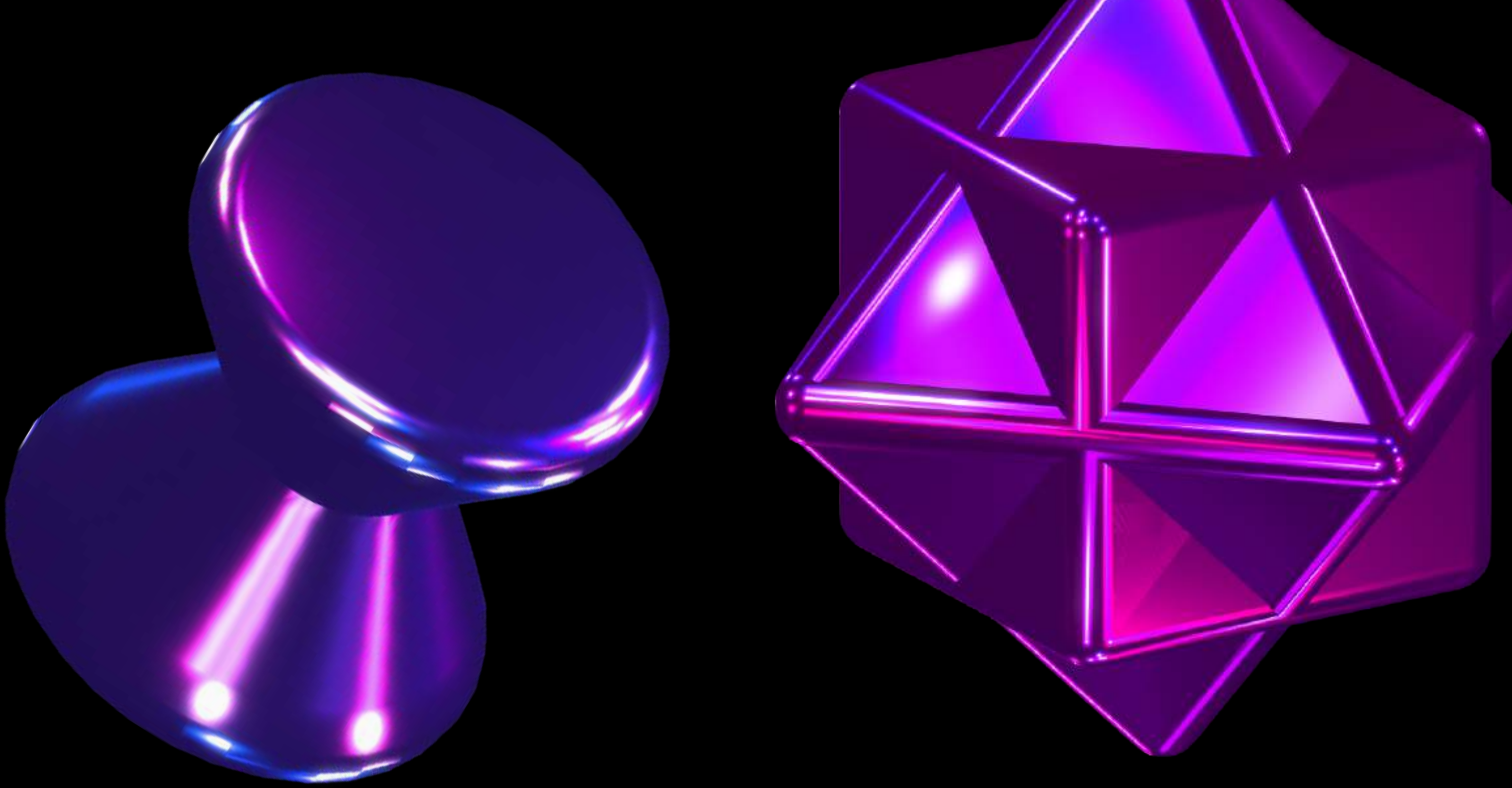


• **FIGURE 8**: ABOUT HALF OF TEAMS USE SYNTHETIC DATA TO ADDRESS INSUFFICIENT EXAMPLES OF EDGE CASES FROM REAL-WORLD DATA. [NOTE: THIS QUESTION WAS MULTI-SELECT]

> The power of simulation and assimilation for us is not just to help us validate things, but to really help us provide an extensive set of synthetic data based on data that we've already collected and use that to dramatically extend into different cases that you normally wouldn't see driving on the road. We can put in different weather conditions, vegetation, and environment around the same segment of road and just create a lot more types of data.

Yangbing Li,

Senior Vice President of Software, Aurora



Model Development & Deployment

Once teams have data, the next steps in the ML lifecycle are model development, deployment, and monitoring. Building a high-performance ML model typically requires iterations on the dataset, data augmentation, comparative testing of multiple model architectures, and testing in production.

Even after a model is deployed, it is important to monitor performance in production. Precision and recall may be relevant to one business, while another might need custom scenario tests to ensure that model failure does not occur in critical circumstances. Tracking these metrics over time to ensure that model drift does not occur is also important, particularly as businesses ship their products to more markets or as changing environmental conditions cause a model to suddenly be out of date. In this chapter, we explore the key challenges ML teams encounter when developing, deploying, and monitoring models, and we discuss some best practices in these areas.

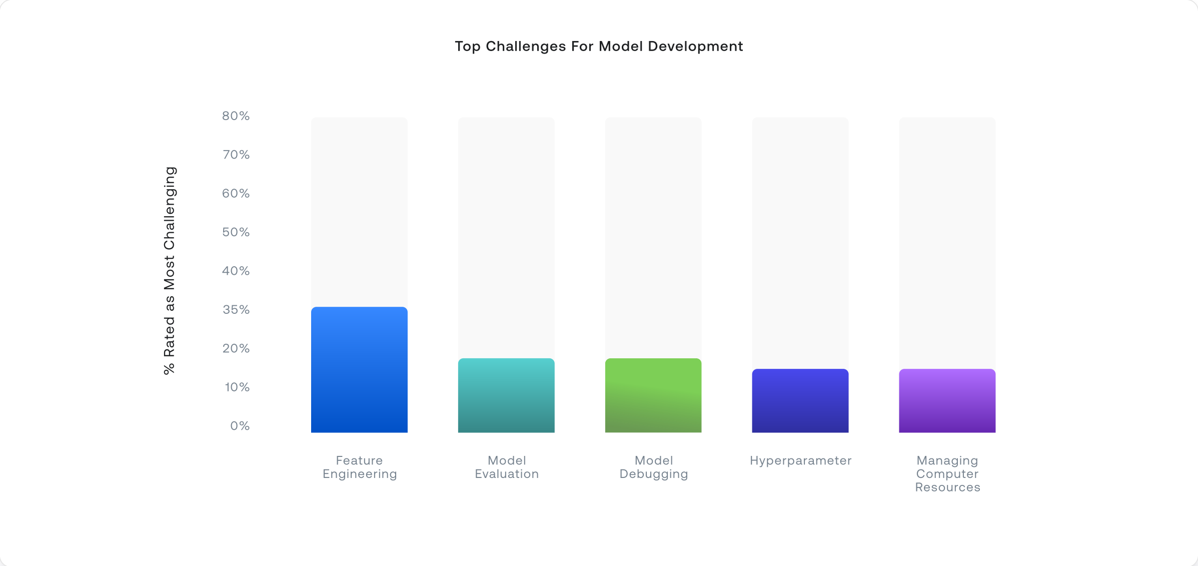
03

Model Development & Deployment Challenges

Feature engineering is the biggest challenge in model development.

The majority of respondents consider feature engineering to be the most challenging aspect of model development. Feature engineering is particularly relevant for creating models on structured data, such as predictive models and recommendation systems, while deep learning computer vision systems usually don't rely on feature engineering.

Challenge 1



• **FIGURE 9**: FEATURE ENGINEERING IS THE BIGGEST CHALLENGE TO MODEL DEVELOPMENT.

For tabular models, however, feature engineering can require several permutations of logarithms, exponentials, or multiplications across columns. It is important to identify co-linearity across columns, including "engineered" ones, and then choose the better signal and discard the less relevant one. In interviews, ML teams expressed the concern of choosing columns that border on personally identifiable information (PII), in some cases more sensitive data leading to a higher-performing model. Yet in some cases, analogs for PII can be engineered from other nonsensitive columns.

We hypothesize that feature engineering is time-consuming and involves significant cross-functional alignment for teams building recommendation systems and other tabular models, while feature engineering remains a relatively foreign concept for those working on deep learning–based vision systems such as autonomous vehicles.

> All the science is infused with engineering, discipline, and careful execution. And all the engineering is infused with scientific ideas. The reason is because the field is becoming mature, so it is hard to just do small-scale tinkering without having a lot of engineering skill and effort to really make something work.

Ilya Sutskever,

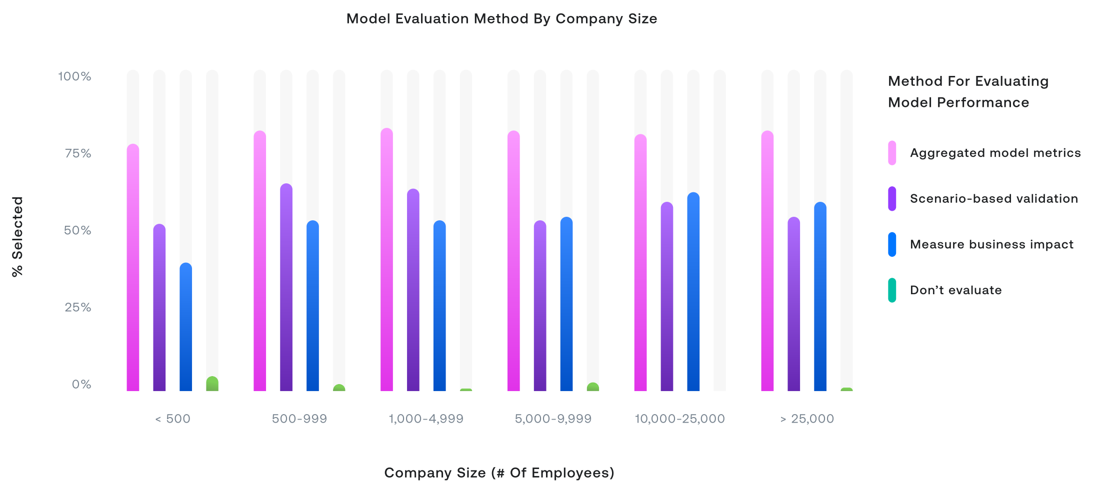
Co-founder and Chief Scientist, OpenAI

Measuring the business impact of models is a challenge, especially for smaller companies.

When asked how they evaluate model performance, a majority (80%) of respondents indicated they use aggregated model metrics, but far fewer do so by performing scenario-based validation (56%) or by evaluating business impact (51%). Our analyses showed that smaller companies (those with fewer than 500 employees) are least likely to evaluate model performance by measuring business impact (only 40% use this approach), while large companies (those with more than 10,000 employees) are most likely to evaluate business impact (about 61% do so).

Challenge 2

We hypothesize that as organizations grow, there is an increasing need to measure and understand the business impact of ML models. The responses also suggest, however, that there is significant room for improvement across organizations of all sizes to move beyond aggregated model metrics into scenario-based evaluation and business impact measurement.



• **FIGURE 10**: SMALL COMPANIES ARE LEAST LIKELY AND LARGE COMPANIES ARE MOST LIKELY TO EVALUATE MODEL PERFORMANCE BY MEASURING BUSINESS IMPACT.

We need to shift to business metrics—for example, cost savings or new revenue streams, customer satisfaction, employee productivity, or any other [metric] that involves bringing the technical departments and the business units together in a combined life cycle, where we measure all the way to the process from design to production. MLOps is all about bringing all the roles that are involved in the same lifecycle from data scientists to business owners. When you scale this approach throughout the company, you can go beyond those narrow use cases and truly transform your business.

David Carmona,
General Manager, Artificial intelligence and Innovation, Microsoft

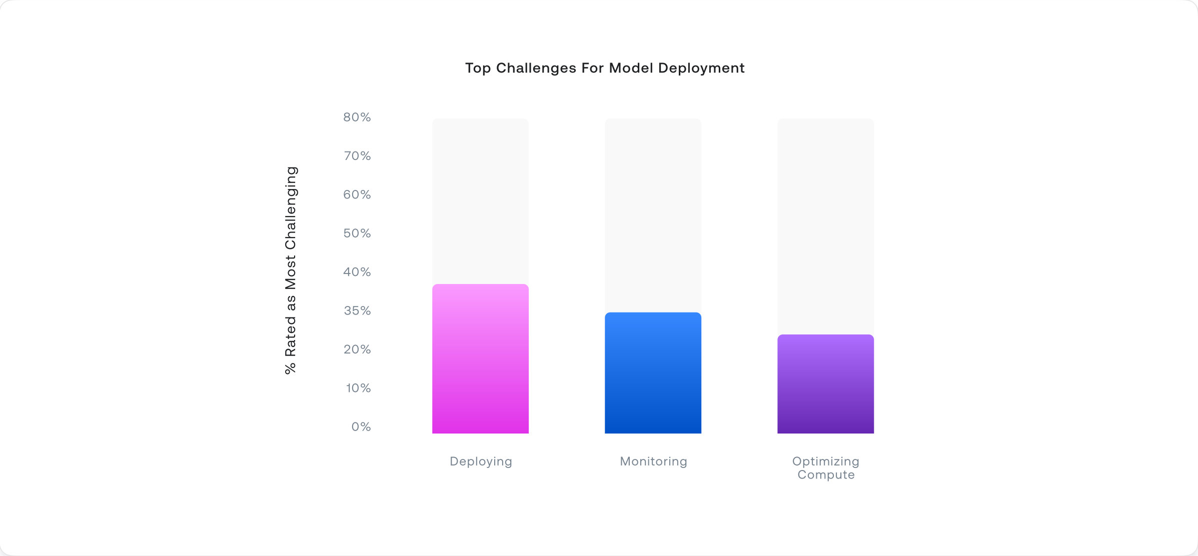
The majority of ML practitioners cite deployment at scale as a challenge.

Deploying was cited by 38% of all respondents as the most challenging part of the deployment and monitoring phase of the ML lifecycle, followed closely by monitoring (34%) and optimizing compute (30%). B2B services rated deploying as most challenging, followed by retail and e-commerce.

Our interviews suggest that deploying models is a relatively rare skill set compared to data engineering, model training, and analytics. One reason: while most recent graduates of undergraduate and postgraduate programs have extensive experience training and tuning models, they have never had to deploy and support them at scale. This is a skill one develops only in the workplace. Typically, at larger organizations, one engineer is responsible for deploying multiple models that span different teams and projects, whereas multiple engineers and data scientists are typically responsible for data and training.

Monitoring model performance in production came in a close second, indicating that, like model deployment, monitoring a distributed set of inference endpoints requires a different skill set than training, tuning, and evaluating models in the experimentation phase. Lastly, optimizing compute was ranked as the third challenge by all groups, which is a testament to the advances in usability of cloud infrastructure, and also the rapid proliferation of expertise in deploying services across cloud providers.

Challenge 3



> A lot of people want to work on the cool AI models, but when it comes down to it, there's just a lot of really hard work in getting AI into a production-level system. There's so much work in labeling the data, cleaning the data, preparing it, compared to standard engineering work and load balancing things and dealing with spikes and so on. A lot of folks want to say they're working on AI, but they don't want to do most of these really hard aspects of creating the AI system.

Richard Socher,
CEO, You.com

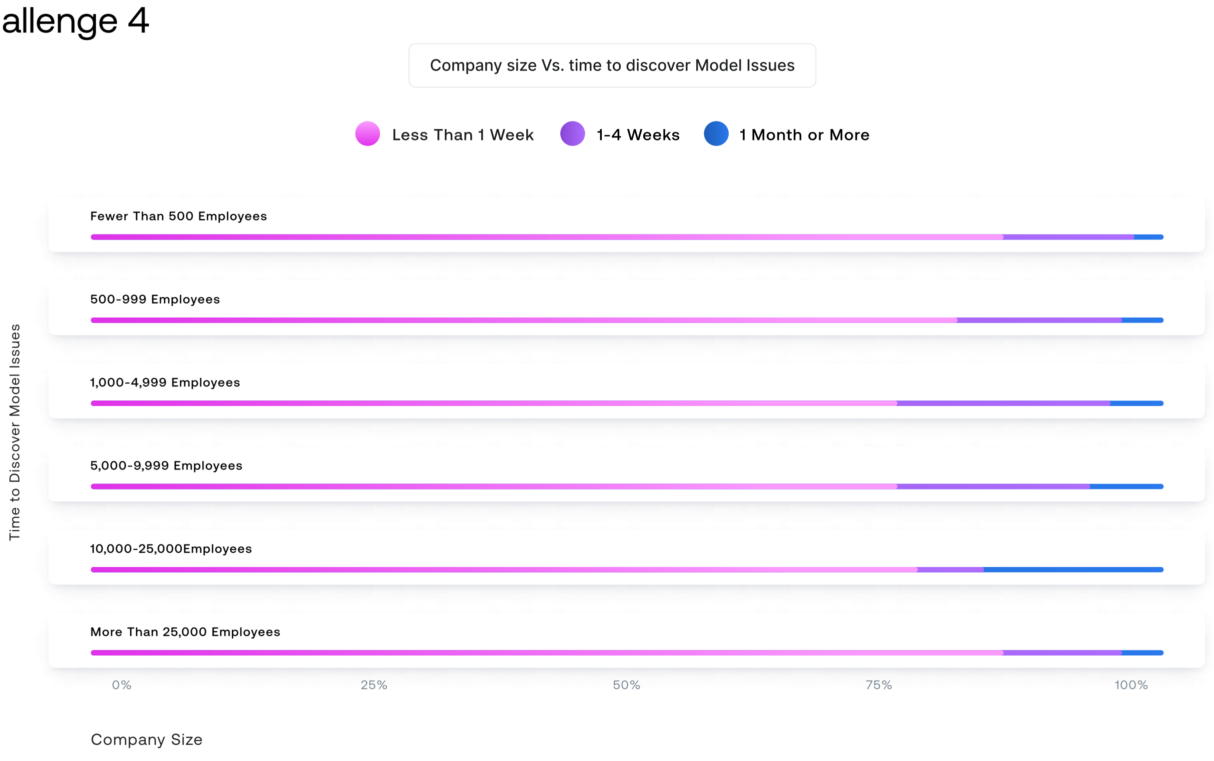
Large companies take longer to identify issues in models.

Smaller organizations can usually identify issues with their models quickly (in less than one week). Larger companies (those with more than 10,000 employees) may be more likely to measure business impact, but they are also more likely than smaller ones to take longer (one month or more) to identify issues with their models.

It is not that larger companies are less capable—in fact, they may be more capable—but they operate complex systems at scale that hide more complex flaws. While a smaller company may try to serve the same model to five hypothetical customers and experience one failure, a larger enterprise might push out a model to tens of thousands of users across the globe. Failures might be regional or prone to multiple circumstantial factors that are challenging to anticipate or even simulate.

Larger ML infrastructure systems also develop technical debt such that even if engineers are monitoring the real-time performance of their models in production, there may be other infrastructure challenges that occur simply due to serving so many customers at the same time. These challenges might obscure erroneous classifications that a smaller team would have caught. For larger businesses, the fundamental question is: can good MLOps practices around observability overcome the burden of operating at scale?

Challenge 4

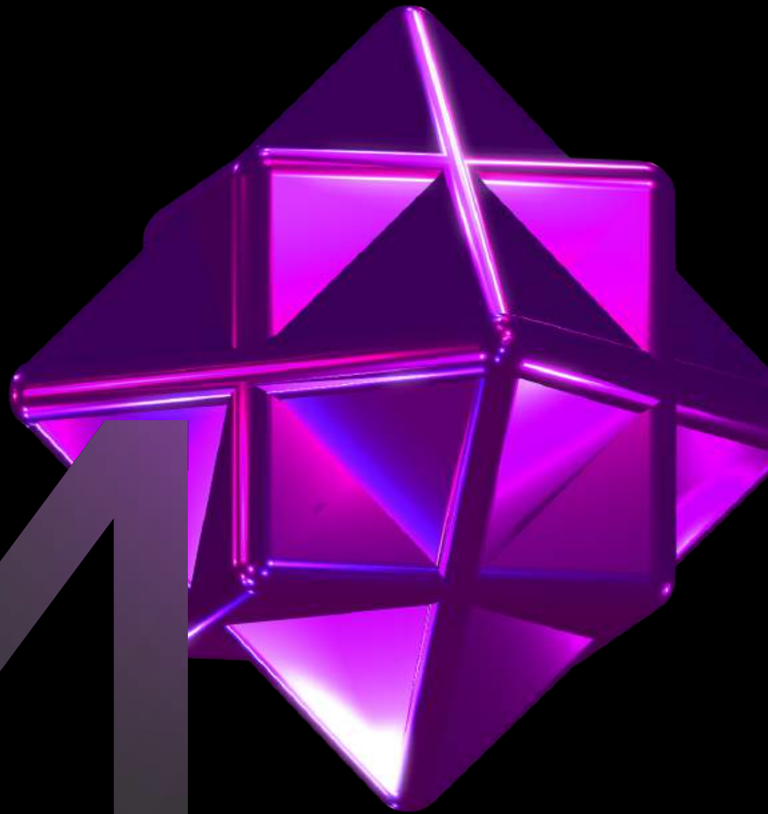


• **FIGURE 12**: SMALLER COMPANIES (THOSE WITH FEWER THAN 500 EMPLOYEES) ARE MORE LIKELY TO IDENTIFY ISSUES IN LESS THAN ONE WEEK. LARGE COMPANIES (WITH 10,000 OR MORE EMPLOYEES) ARE MORE LIKELY TO TAKE MORE THAN ONE MONTH TO IDENTIFY ISSUES.

> I've been at companies that are much smaller in size and companies that are way bigger. Silos get bigger as the companies get bigger. What happens oftentimes at big companies is [we have] introductions to who our users are going to be, the stakeholders, and their requirements. Then off we go in a silo, build a solution, come back six months later, and now the target has moved because all of us are undergoing quick transformations. It's important to involve users at every aspect of the journey and establish those checkpoints, just as we would do for any other external-facing users.

Deepna Devkar,
Vice President of Machine Learning and
Data Platform Engineering, CNN

04



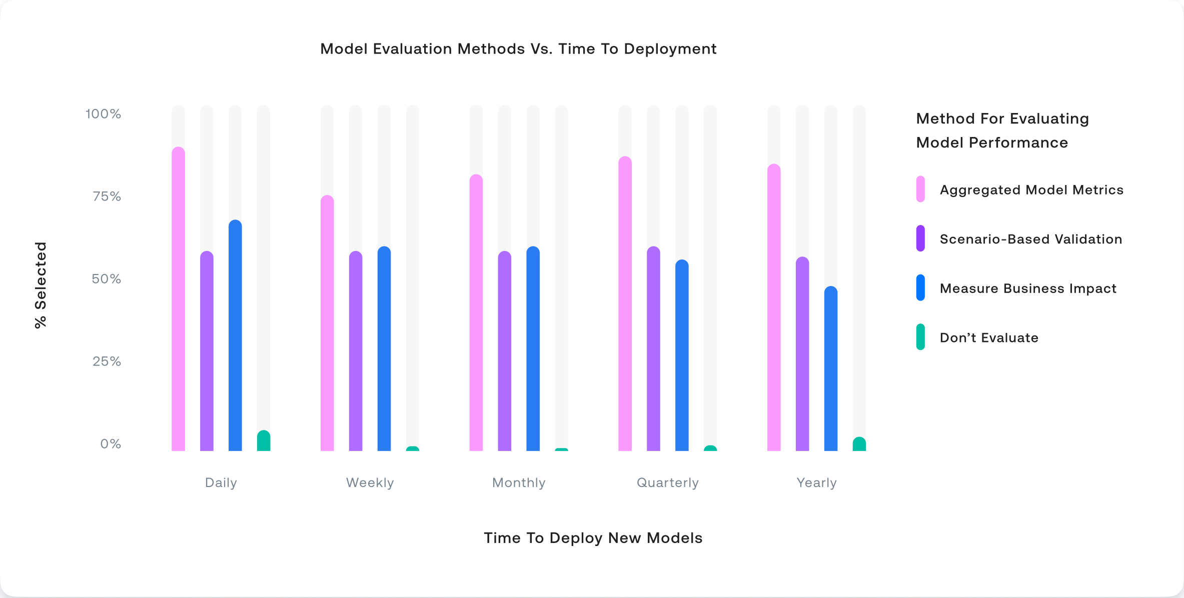
Model Development and Deployment Best Practices

Organizations that retrain existing models more frequently are most likely to measure impact and aggregated model metrics to evaluate model performance.

More frequent retraining of existing models is associated with measuring business impact and using aggregate model metrics to evaluate model performance.

Teams that retrain daily are more likely than those that retrain less frequently to use aggregated model metrics (88%) and to measure business impact (67%) to evaluate model performance. Our survey also found that the software/ Internet/telecommunications, financial services, and retail/e-commerce industries are the most likely to measure model performance by business impact.

Best Practice 1



• **FIGURE 13**: COMPANIES THAT RETRAIN DAILY ARE MORE LIKELY THAN THOSE THAT RETRAIN LESS FREQUENTLY TO USE AGGREGATED MODEL METRICS (88%) AND BUSINESS IMPACT MEASUREMENT (67%) TO EVALUATE THEIR MODEL PERFORMANCE.

> When we confuse the research side, which is developing cooler and better algorithms, with the applied side, which is serving amazing models—predictions at scale to users—that's when things go wrong. If you start hiring the wrong kind of worker, you build out your applied machine learning team only with the Ph.D.'s who develop algorithms, so what could possibly go wrong? Understanding what you're selling, understanding whether what you're selling is general-purpose tools for other people to use or building solutions and selling the solutions—that's really, really important.

Cassie Kozyrkov,
Chief Decision Scientist, Google

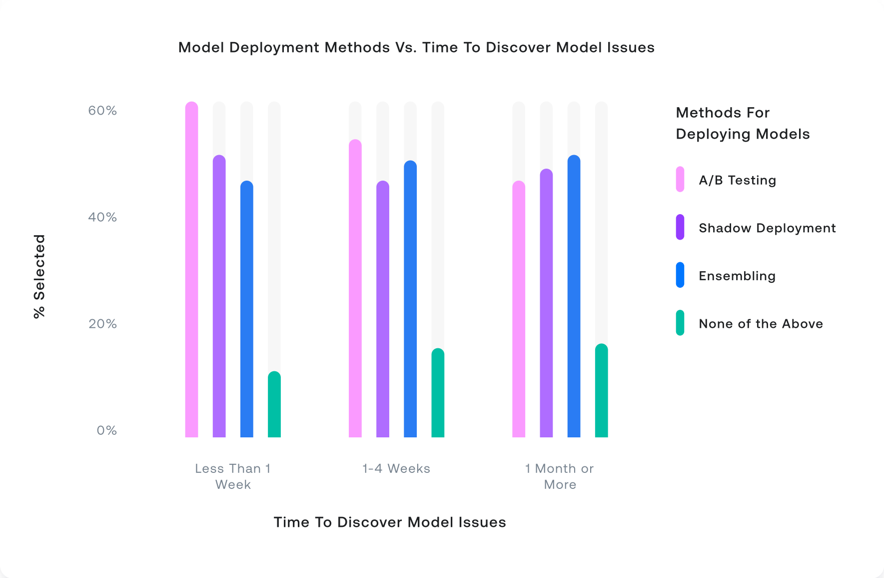
ML teams that identify issues fastest with their models are most likely to use A/B testing when deploying models.

Aggregate metrics are a useful baseline, but as enterprises develop a more robust modeling practice, tools such as "shadow deployments," ensembling, A/B testing, and even scenario tests can help validate models in challenging edge-case scenarios or rare classes. Here we provide the following definitions:

- A/B testing: comparing one model to another in training, evaluation, or production. Model A and Model B might differ in architecture, training dataset size, hyperparameters, or some other factor.
- Shadow deployment: the deployment of two different models simultaneously, where one delivers results to the customer and the developers, and the second only delivers results to the developers for evaluation and comparison.
- Ensembling: the deployment of multiple models in an "ensemble," often combined through conditionals or performance-based weighting.

Although small, agile teams at smaller companies may find failure modes, problems in their models, or problems in their data earlier than teams at large enterprises, their validation, testing, and deployment strategies are typically less sophisticated. Thus, with simpler models solving more uniform problems for customers and clients, it's easier to spot failures. When the system grows to include a large market or even a large engineering staff, complexity and technical debt begin to grow. At scale, scenario tests become essential, and even then it may take longer to detect issues in a more complex system.

Best Practice 2



• **FIGURE 14**: TEAMS THAT TAKE LESS THAN ONE WEEK TO DISCOVER MODEL ISSUES ARE MORE LIKELY THAN THOSE THAT TAKE LONGER TO USE A/B TESTING FOR DEPLOYING MODELS.

> I became pretty open to considering A/B testing, king-of-the-hill testing, with AI-based algorithms. Because we got scalability, we got faster decision making. All of a sudden, there were systems that always had humans in the loop, humans doing repetitive tasks that didn't really need humans in the loop anymore.

Jeff Wilke,

Former CEO, Worldwide Consumer, Amazon

Chairman and Co-Founder of Re:Build Manufacturing

Conclusion

Seismic shift in the rate of artificial intelligence (AI) adoption

The Scale Zeitgeist AI Readiness Report 2022 tracks the latest trends in AI and ML readiness. Perhaps the biggest AI readiness trend has been the seismic shift in the rate of AI adoption. Moving beyond evaluating use cases, teams are now focused on retraining models and understanding where models fail. To realize the true value of ML, ML practitioners are digging deep to operationalize processes and connect model development to business outcomes.

This artificial intelligence report explored the top challenges ML practitioners face at every stage of the ML lifecycle as well as best practices to overcome some of these challenges to AI readiness. We found that most teams, regardless of industry or level of AI advancement, face similar challenges. To combat these challenges, teams are investing in data infrastructure and working closely with annotation partners. We also explored different tactics for model evaluation and deployment. Although deployment methods and tools vary widely across businesses, teams are evaluating business metrics, using A/B testing, and aggregating model metrics. Wherever ML teams are in their development, we can share knowledge and insights to move the entire industry forward.

Our goal at Scale AI is to accelerate the development of AI applications. To that end, this report has showcased the state of AI readiness, identifying the key AI trends in 2022 in terms of the challenges, methods, and successful patterns used by ML practitioners who actively contribute to ML projects.

"AI is a very powerful technology, and it can have all kinds of applications. It's important to prioritize applications that are exciting, that are solving real problems, that are the kind of applications that improve people's lives, and to work on those as much as possible."

Ilya Sutskever — Co-founder and Chief Scientist, OpenAI

Methodology

This survey was conducted online within the United States by Scale AI from March 31, 2022, to April 12, 2022. We received 2,142 total responses from ML practitioners (e.g., ML engineers, data scientists, development operations, etc.). After data cleaning and filtering out those who indicated they are not involved with AI or ML projects and/or are not familiar with any steps of the ML development lifecycle, the dataset consisted of 1,374 respondents. We examined the data as follows:

The entire sample of 1,374 respondents consisted primarily of data scientists (24%), ML engineers (22%), ML researchers (16%), and software engineers (13%). When asked to describe their level of seniority in their organizations, nearly half of the respondents (48%) reported they are an individual contributor, nearly one-third (31%) said they function as a team lead, and 18% are a department head or executive. Most come from small companies with fewer than 500 employees (38%) or large companies with more than 25,000 employees (29%). Nearly one-third (32%) represent the software/Internet/telecommunications industry, followed by healthcare and life sciences (11%), the public sector (9%), business and customer services (9%), manufacturing and robotics (9%), financial services (9%), automotive (7%), retail and e-commerce (6%), media/entertainment/hospitality (4%), and other (5%).

When asked what types of ML systems they work on, nearly half of respondents selected computer vision (48%) and natural language processing (48%), followed by recommendation systems (37%), sentiment analysis (22%), speech recognition (10%), anomaly detection/classification/reinforcement learning/predictive analytics (5%), and other (17%).

Most respondents (40%) represent organizations that are advanced in terms of their AI/ML adoption—they have multiple models deployed to production and regularly retrained. Just over one-quarter (28%) are slightly less advanced—they have multiple models deployed to production—while 8% have only one model deployed to production, 14% are developing their first model, and 11% are only evaluating use cases.

Group differences were analyzed and considered to be significant if they had a p-value of less than or equal to 0.05 (i.e., 95% level of confidence).

About Scale

Scale AI builds infrastructure for the most ambitious artificial intelligence projects in the world. Scale addresses the challenges of developing AI systems by focusing on the data, the foundation of all AI applications. Scale provides a single platform to manage the entire ML lifecycle from dataset selection, data management, and data annotation, to model development.

https://scale.com/