Labelbox

10 training data practices used by the most successful AI teams

00 Content

01 Introduction

At Labelbox, we're in a position to work with an incredibly wide variety of companies all intent on building something remarkable with AI and neural networks.

We've all heard about the transformative power of machine learning, but we don't always get a chance to hear about the hard work it takes to get to production AI. Nearly 96% of companies encounter delays getting to production, and 78% of machine learning projects stall before deployment.[1]

Having helped hundreds of organizations across a wide range of verticals, we don't always get full visibility into everyone's model outcomes, but we have discovered some commonalities in their ML workflows.

We've been able to identify some key roadblocks and some shared best practices of highly successful ML teams and how they've utilized certain processes, standards, and tools to more skillfully master how they work with training data.

This guide will lay out ten training data practices used by the most successful AI teams as well as highlight a few customer success stories where these practices were applied in production use cases. We hope you can incorporate these lessons into your own MLOps and accelerate your AI journey.

[1] *Algorithmia 2020 State of ML Survey*

02 Defining and structuring high-quality training data

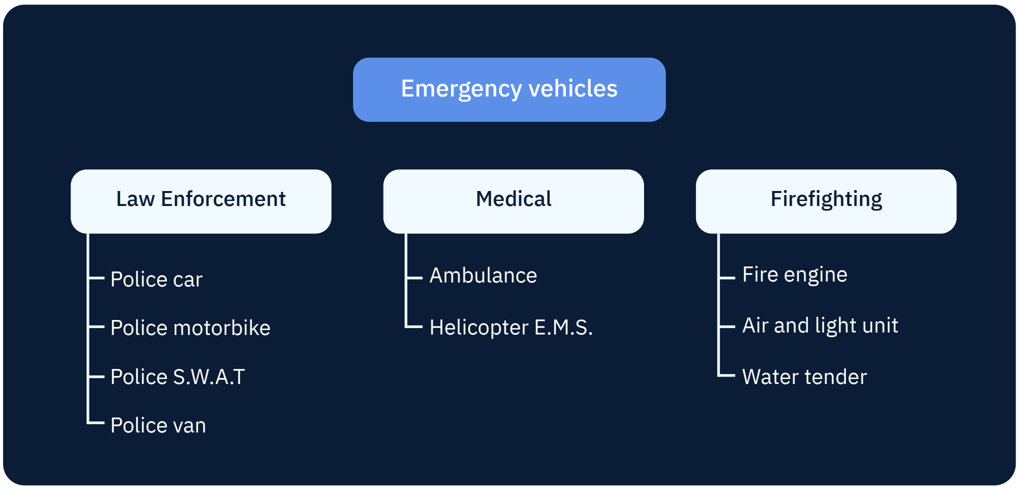
In his popular blog article on training data and neural networks, Andrej Karpathy, director of artificial intelligence and Autopilot Vision at Tesla, said: “Become one with the data.”[2]

We’ve observed that a key belief held by successful ML teams is this: **high-quality training data produces the most performant models**.

As you annotate your data, it is important that you clearly define **what** it is you want to be labeled and **how** you want it labeled. An ontology does just that: it gives your labelers a structured definition of what output you want from your labeling process. A well-designed information architecture is critical in order to support your AI initiatives and provide meaningful structure to your data.

By doing so, an ontology ensures that all the labels in a project follow the same set of rules, reducing the possibility of confusion for your ML/AI model. In addition, ontologies provide a reusable, adaptive structure for teams that, in essence, serve as your intellectual property (IP) and should be carefully curated and updated. We’ve learned from leading AI teams that their ability to easily reuse and maintain ontologies without errors is critical to their AI development’s success.

For example, in the case of emergency vehicle detection, is it enough to group all emergency vehicles in a single category? If you want to build different rules, you may want to add additional categories to differentiate police cars from ambulances. Some regions may use police motorbikes, giving rise to a potential new topic or category. Ontologies are not immutable – they are living documents – so the use and intent of language are constantly in flux.



[2] *A Recipe for Training Neural Networks by Andrej Karpathy*

A clean, thoughtful ontology is critical for creating high-quality labeled data with minimal errors and inconsistencies. Ontologies are an essential part of the Labelbox labeling platform.

As a best practice, leading AI teams now employ specific tooling to make ontology management easier, which includes the ability to:

- Create ontologies from scratch, reuse ontologies from other labeling projects, and clone ontologies
- Query for specific labels within all projects that use the same ontology
- Manage their ontology via the user interface or API

Enterprises may have hundreds of AI projects, so enabling their ontology to scale is essential. As opposed to a series of disjointed projects, a repeatable framework can be designed to reduce lag and confusion. With a strong ontology in place, adjustments can be made to the data in one location and propagate through the existing associative relationships.

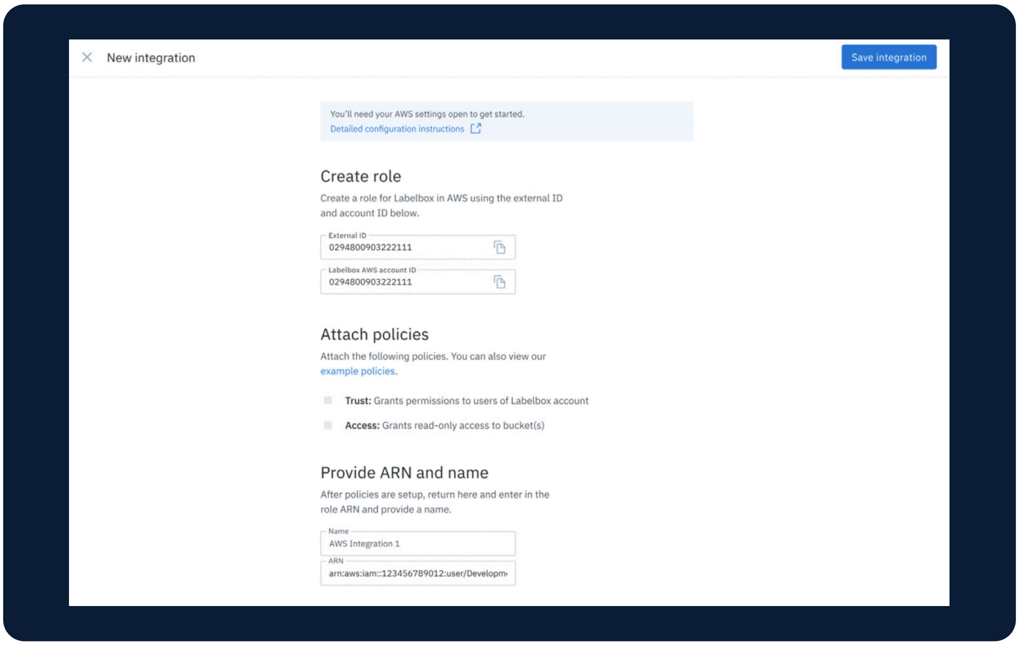
03 Leveraging APIs for smooth integrations and delegated access for better access to data

Easier ways to access and manage data is another key tenet of successful ML teams. We've seen that this is typically accomplished by tapping into the most convenient ways to import, store, and export data.

For example, leading AI teams leverage APIs for smooth integrations, which allows them to focus on analyzing fresh data to stimulate innovation and promote experimentation, rather than spending time and energy managing training data infrastructure. An API-first approach provides ML teams with a stream of integrated and clean data, which creates better insights and more data-driven decisions.

We've seen that by utilizing APIs, teams can flexibly integrate training data into their existing data collection, storage, and management systems. A complete API can automate the data import and labeled data export processes, all while retaining the metadata of how a specific dataset was created, labeled, as well as any labeler feedback.

To complement the use of APIs, delegated access also helps with data management, especially in the realm of access control. With delegated access, you can control who in your organization is authenticated and authorized to access the raw data stored in your cloud storage. Organizations struggling with data silos and little to no asset sharing can move to discovering, sharing, and reusing data and assets. In addition, delegated access allows you to securely and seamlessly host your labeling assets while providing external applications with the limited access they need. With this approach, you can more easily store your assets in buckets from a cloud provider of your choice and use native Identity and Access Management (IAM) roles and policies to control access.



You can quickly set up, validate, and manage your integrations in one place, shown here with Labelbox and AWS storage as an example.

By continuously monitoring and improving your IAM, you'll protect your data and IP from unauthorized access. This includes consistent audits of access logs and admin activity for more effective data governance and security. As we established earlier, your training data is your IP and leading AI teams get ahead of any potential threats by creating timely alerts and mitigating any unauthorized access.

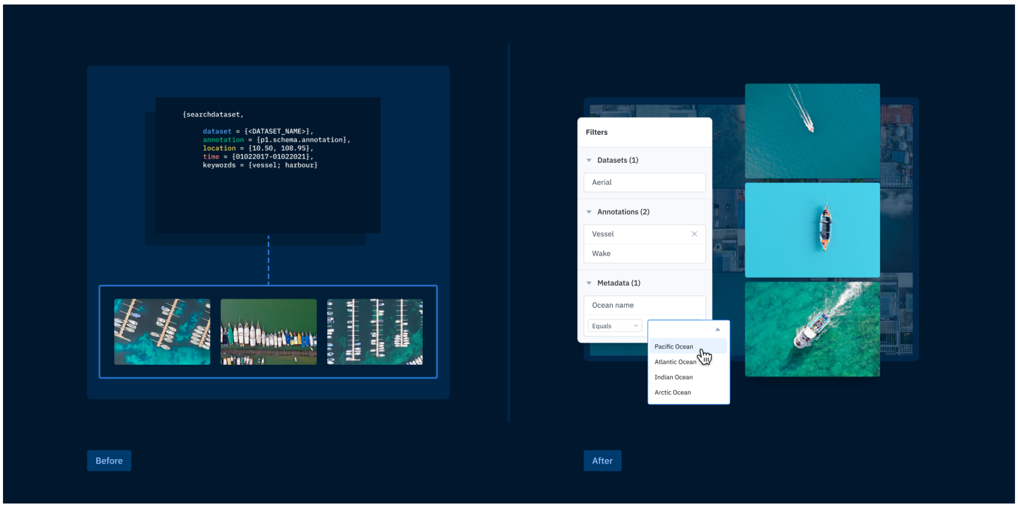
04 Interpreting model performance and prioritizing areas for iteration

There is a reason for every decision your model makes, and those decisions typically begin and end with your data.

When it comes to enabling production AI, understanding your model is key. And when it comes to model performance, the overall performance of your model will rarely give you the insight you need into how your model is achieving specific predictions. The only way to truly understand the inner workings of your model is to visualize and interact hands-on with your data and do so in an efficient way.

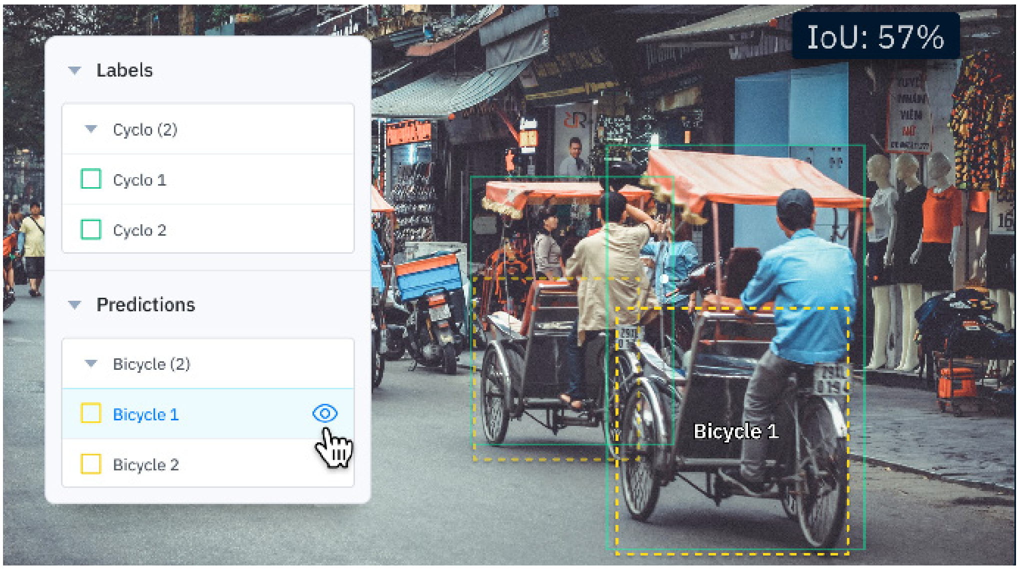
For example, in the case of object detection, you may be visualizing magnitudes of bounding boxes across a wide range of classifications. This is frequently done by relying on one-off query scripts, which can be very time-intensive and quite inflexible once the images are regenerated.

Fortunately, the market is starting to catch up with this problem and there has been an emergence of tools to help organizations visualize model errors, explore their datasets, and take action much more easily.



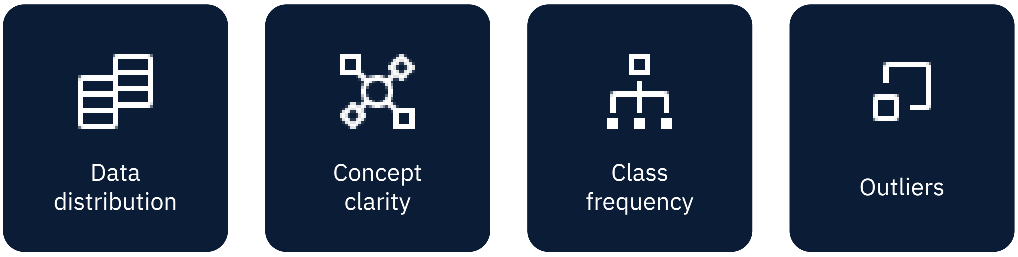
Frequently, ML engineers need to write one-off queries to find data, as conceptualized on the left. This inefficient process can slow down your MLOps. But now, teams are adopting flexible and scalable code-free search tools, as visualized on the right in Labelbox.

As an example of one specific workflow being adopted, leading AI teams are now identifying trends in model behavior by slicing and selecting data to surface patterns in performance across different cohorts of training data. Afterward, they can sort and filter by heuristics like IoU (Intersection over Union) and confidence, as well as annotation classes and custom metadata fields.



Interact with your data visually and compare model predictions to ground truth, shown here in Labelbox.

By better understanding your model, you can prioritize where to focus your next iteration and waste less time with data that won't lift performance. We've heard from our customers that they commonly run into four types of model errors, along with where they prioritize areas for iteration:



Data distribution errors occur when the model is confused by slight differences. To resolve this type of error, teams can sample from a new data stream to determine how much corrective action they need to take.

Example: Data distribution could occur when introducing a new satellite that has slightly different color spectra and resolutions within the same class. This can affect the model and result in corner case effects.

Concept clarity issues arise when new classifications, incorrect labels, and/or incomplete ontologies cause the model performance to suffer. It can be addressed by optimizing labeling operations and folding domain expertise into the process.

Example: A series of boats are docked but the team didn't label individual sailboats since they weren't related to their concept of vessel detection. However, this particular case may cause confusion for labelers or the model. Having purer classes and more complete and concise ontologies can enhance model performance.

Class frequency errors occur when the model hasn't been trained enough on a particular class to reach an acceptable performance level. There are not enough instances for the model to learn a pattern. Through methods like leveraging active learning and using similarity search tools to mine for more assets of a particular class, teams can correct the large imbalances between class representation within the training data.

Example: A model was built to detect roofs. Later on, it was discovered that the model didn't perform particularly well on mansions. Due to a class imbalance, there were not enough examples of mansions in the dataset, but this issue wasn't recognized until the ML team performed diagnostics on the model. The team previously spent time and effort adding many tight residential blocks to the dataset, unaware that the model wouldn't see much lift with this data. With better model diagnostics processes and tooling earlier on, the team would have known where to prioritize the iteration and could have focused on adding more mansions to the dataset to boost performance.

Outlier errors are objects that deviate drastically from the rest of the other objects. Outliers are unlikely observations in a dataset and can have a wide range of causes.

Outliers are hard to define in nature because of the specifics of each dataset. They can be addressed by either excluding that type of data entirely from the dataset (although this must be done cautiously), rebuilding the ontology, or by redesigning the approach to include that circumstance.

Even though these are just four illustrative classes of errors, there will always be a range of issues that occur before and after a model is in production. Having reliable and robust model diagnostics tooling is a critical part of ML success and scaling your ML efforts.

Labelbox customer story: CAPE Analytics

CAPE Analytics is a leading provider of AI-powered geospatial property data. The company is building a best-in-class training data pipeline and heavily employing APIs. CAPE uses deep learning and geospatial imagery to provide instant property intelligence for buildings across the United States. Their pioneering work for assessing risk and property value is tied to common property factors (pools, roof construction, solar panels) but also emerging risks like vegetation coverage, detailed roof condition, and wildfire zones.

“We're going through climate change right now. And so it becomes even more important to accurately predict the impact of wildfires and hurricanes and other natural disasters. AI is going to play a central role in all those things. Deep learning models are very, very data-hungry,” said Sheetanshu Pandey, Chief Technology Officer at CAPE Analytics.

“So you can always collect as much data, as many labels as you want, and you can keep improving the quality of the models. And really, the challenge here that CAPE has been able to overcome is finding that right balance.”

To accomplish this feat, CAPE’s engineering and data science team searched extensively for ways to build a complete and automated training data solution.

Part of this complete solution included automating the data import and labeled data export process, which saved the CAPE team valuable time not spent handling data transfer methods manually. By connecting their data via API, CAPE’s labeled data possessed all the metadata of how a specific dataset was created and labeled, as well as labeler feedback associated with the assets.

CAPE Analytics’ engineering team programmatically customized Labelbox’s API experience so that they could interrupt the cycle within the training data process and grab data out, if needed, to fit seamlessly with how their ML workflow was structured.

This made the process much easier to manage and track as the complexity of their projects grew over time. For comparison, CAPE’s previous homegrown tools lacked this type of functionality and relied more on manual and labor-intensive workflows. To put all these time savings into perspective, Labelbox took what used to take several days (in order to set up a project and an additional day to fetch the data) and turned it into a simple 10-minute process.

“We’ve seen AI and machine learning dramatically change the insurance sector,” said Kayvan Farzaneh, Head of Marketing at CAPE Analytics. “People really want a high-quality customer experience and speed. And so insurance has had to adapt to that world as well and be able to provide quick quotes, really accurate underwriting and pricing of those insurance policies. They need really high quality, accurate and recent information in order to do that.”

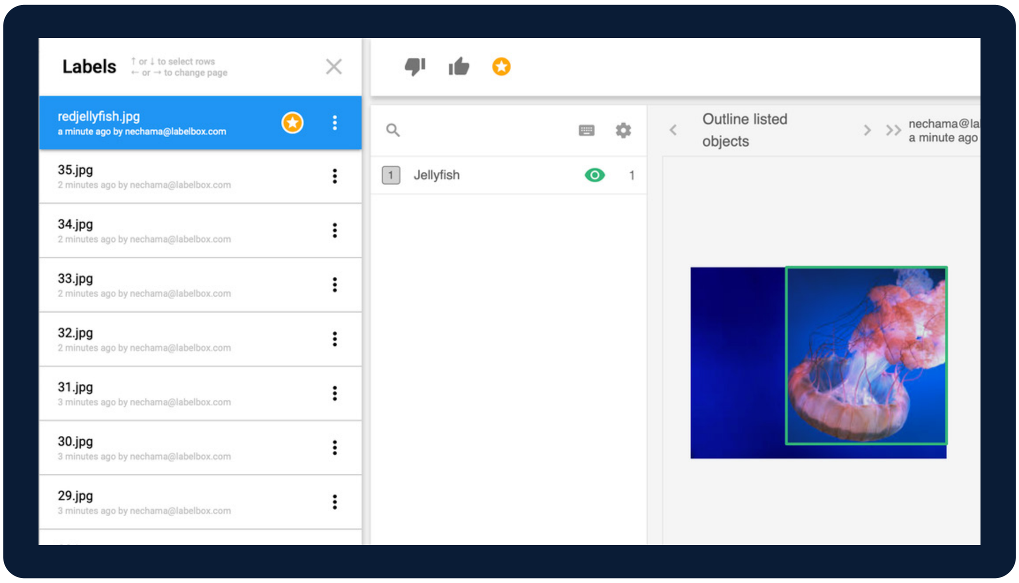
Watch the CAPE Analytics Breakthrough Customer of the Year Video to hear how the CAPE team unraveled their data and streamlined their MLOps processes.

05 Creating benchmark standards

As we stated earlier, we've seen that the most successful ML teams typically hone in on their training data.

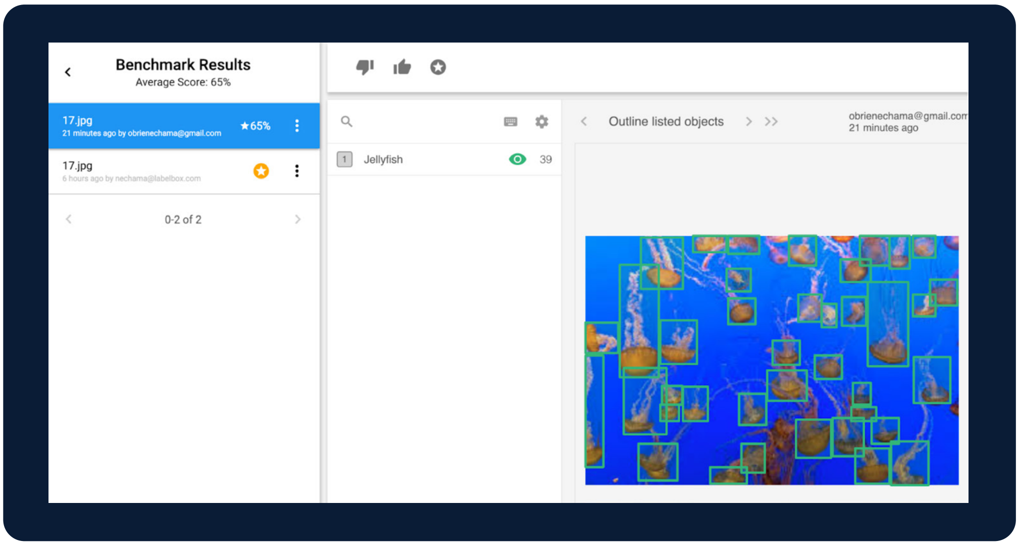
Iterating on training data ensures that the labels are accurate, addresses areas of error or low confidence in the model, and is critical to producing a performant model. These teams have well-defined ground truth and have put measures in place to check that strict standards of quality are being met.

While training data quality is critical for machine learning model performance, quality assurance (QA) is an often overlooked aspect of data labeling operations that can provide tremendous value and insights, particularly when applied to model testing and validation.



A benchmark standard has been set, or a "ground truth" label that all subsequent labels for a given asset can be checked against, shown here in Labelbox and marked with a gold star.

A useful quality assurance practice used by leading AI teams is creating benchmarks for QA. Domain experts or data scientists can set benchmark standards or a "ground truth" label that all subsequent labels for a given asset can be checked against. A benchmark then calculates a numerical value for accuracy by measuring how close a label is to the benchmark. Today, leading AI teams utilize benchmarks in advanced ways such as automatically distributing assets to labelers in order to evaluate their understanding of the labeling task and their accuracy in performing it. In addition, benchmarks allow you to track accuracy by overall project, by labelers, and by labels.



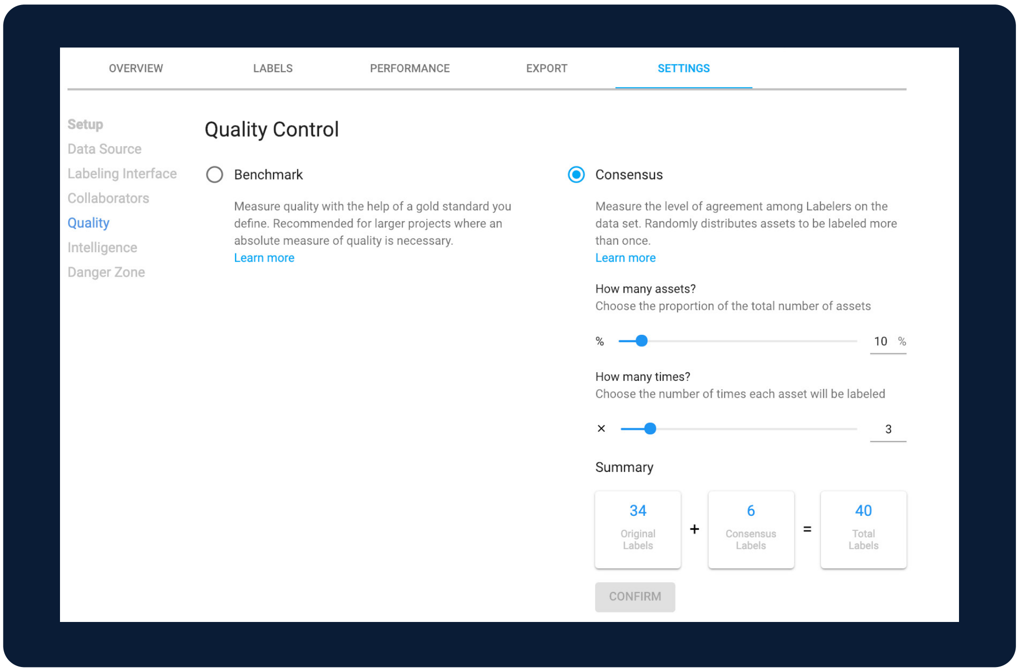
Comparing a crowded jellyfish image to the previously defined benchmark image. While they are very similar, the labels disagree slightly on how much of the tentacle to include in each bounding box.

In our benchmark image example above, it made sense to label the entire body of the jellyfish. Now upon further review, the ML team would need to decide whether it's helpful to label the long tentacles of the jellyfish. Do they confuse the model? Would the model start identifying any comparable wispy shape, like a piece of seaweed, as a jellyfish? If the tentacle is crucial to labeling the jellyfish, the team would likely need to collect and label more instances of long tentacles so the model can become adept at identifying them. The ability to visually monitor and iterate on training data is often overlooked as a massive area of leverage when orienting your labeling operations around quality. This benchmark reviewing process can help you to understand weaknesses in labeling efforts and also reveal potentially problematic ontology configurations.

06 Utilizing consensus scoring to surface areas of low confidence

Another way to improve your training datasets, boost reliability, and orient your labeling operations around quality is through consensus scoring.

Consensus scoring allows you to (1) automatically compare the annotations on a given asset to all other annotations on that asset and (2) measure the rate of agreement between multiple annotators (human or machine).



Consensus scoring in Labelbox

A consensus score is calculated by dividing the sum of agreeing labels by the total number of labels per asset. You can configure the percentage of training data and the number of labelers to test. Consensus scoring can also work to counter any errors or bias of a single labeler or model.

We've found that the best use of consensus is when it is happening in near real-time so you can take immediate and corrective actions towards improving your training data and model performance.

07 Facilitating collaboration between domain experts, labelers, and data scientists

While smooth communication and collaboration may seem obvious to the success of any project at any level, it's almost always one of the quickest pitfalls. When it comes to machine learning operations, there can often be limited collaboration between domain experts, labelers, and data scientists, and no coherent plan for making decisions affecting the team's ML initiatives and goals. Many teams experience bottlenecks during the machine learning process due to misalignment or miscommunications and that sets up a key roadblock often seen in stalled AI projects.

In many cases, team members are sharing specific issues with their colleagues via email or chat, but referencing back to a particular dataset or edge case on cloud-based spreadsheets, which can require a lot of hopping between platforms and manual work to get to that specific issue. Without a centralized process and place of record, valuable insights and key issues are more likely to get buried in past communications, become harder to track, and become invisible to other team members.

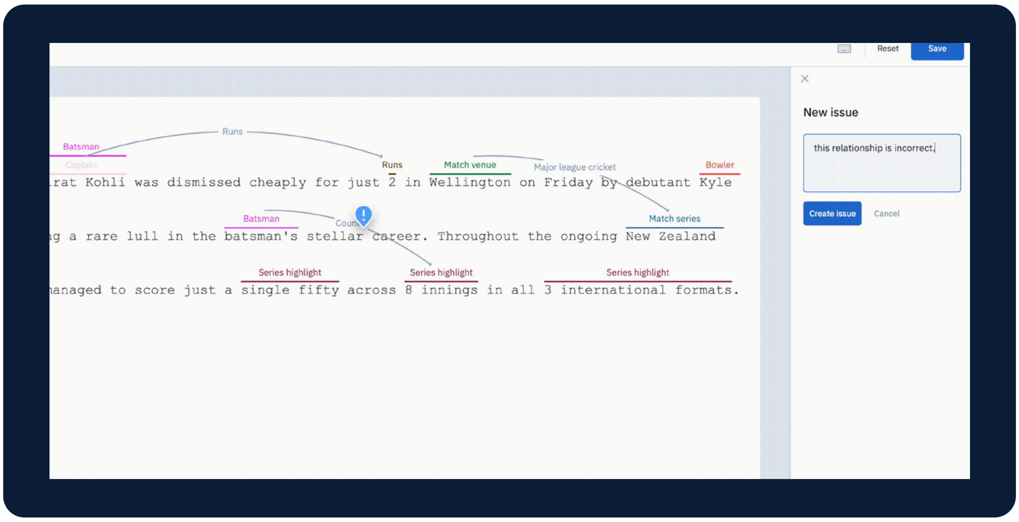
Teams should build a centralized coordination system to achieve strategic goals such as:

- Weighing in on new data collection
- Developing potential new use cases
- Creating benchmarks for ground truth
- Participating in consensus voting
- Clarifying feedback
- Discussing edge cases
- Protecting and improving data quality
- Agreeing on areas for iteration
- Building and maintaining MLOps tools, processes, and infrastructure

Weighing in on new data collection	Developing potential new use cases	Creating benchmarks for ground truth
Protecting and improving data quality	Discussing edge cases	Participating in consensus voting
Agreeing on areas for iteration	Building and maintaining MLOps tools, processes, and infrastructure	Clarifying feedback

Having a way to centralize communication allows for a cross-functional exchange of experiences and ideas that will affect the team's AI efforts and objectives. As a specific example of a tool that is often used by leading AI teams - labelers, reviewers, and other teammates involved in creating and reviewing training data should be able to start conversation threads in their annotation editor while labeling or reviewing labels.

With an "Issues and Comments" tool, labelers can better handle the uncertainty that comes with labeling, especially when they're onboarding or working with particularly complex or subjective data. This tool creates a simple and reliable channel for escalating questions to reviewers or subject matter experts.



Issues and Comments in Labelbox.

A labeler can create an issue to ask a question, submit the label completed to the best of their ability, and receive feedback and clarification during the label review process. Furthermore, labelers and domain experts can create, comment on, and resolve any issue on a label they created while project reviewers and admins have access to issues on any label in their project. With a consolidated view of all this activity, teams can resolve labeling issues faster than ever.

Labelbox customer story: ImageBiopsy Lab

ImageBiopsy Lab builds AI applications to help physicians and the medical community at large better understand and diagnose musculoskeletal diseases (MSK), which affect 1.7 billion people worldwide. They heavily utilize collaboration for the creation of training data. Today, the team focuses on optimizing the radiological workflow via automation and standardization to enable the early detection and prevention of MSK. This automated decision support allows orthopedists and radiologists to assess and predict bone health by turning images into actionable information.

“We use it to connect the dots, discover new correlations between the image and the patient outcome. This will fundamentally change the way healthcare will be done in the future,” says ImageBiopsy Lab Co-founder and COO Christoph Götz.

Managing the flow of immense medical data through their pipeline is not an easy task.

“We usually receive our medical data as disorganized batches, which we must first pre-sort in order to get a basic knowledge of its content. From this, we compile training and validation datasets tailored to address the disease symptoms of a specific patient journey. Those are the ones which will be uploaded to Labelbox for a fine-grained annotation. After the labeling process and a thorough review, the labels are downloaded and stored in our database for additional usage," says Götz.

The team at ImageBiopsy Lab faced challenges when trying to create efficient labeling operations, particularly around managing outsourcers and writing understandable requirements for annotations. They had to ensure their labels fit their regulatory standards as well as their own internal standards. Finding a better way of communicating and collaborating with the disparate teams was a top priority.

"Implants are very diverse and often very rare in the image datasets. Finding common rules for annotation requirements and explaining every edge case to outsourcers was a main hurdle. Luckily the issues & comments function by Labelbox helps to quickly clarify questions on edge cases and make these images available for AI training. This directly translates to more patients benefiting from our solutions," says Götz.

Watch the ImageBiopsy Lab Breakthrough Customer of the Year Video to hear more on how the team at ImageBiopsy Lab streamlined their MLOps to better serve their patients.

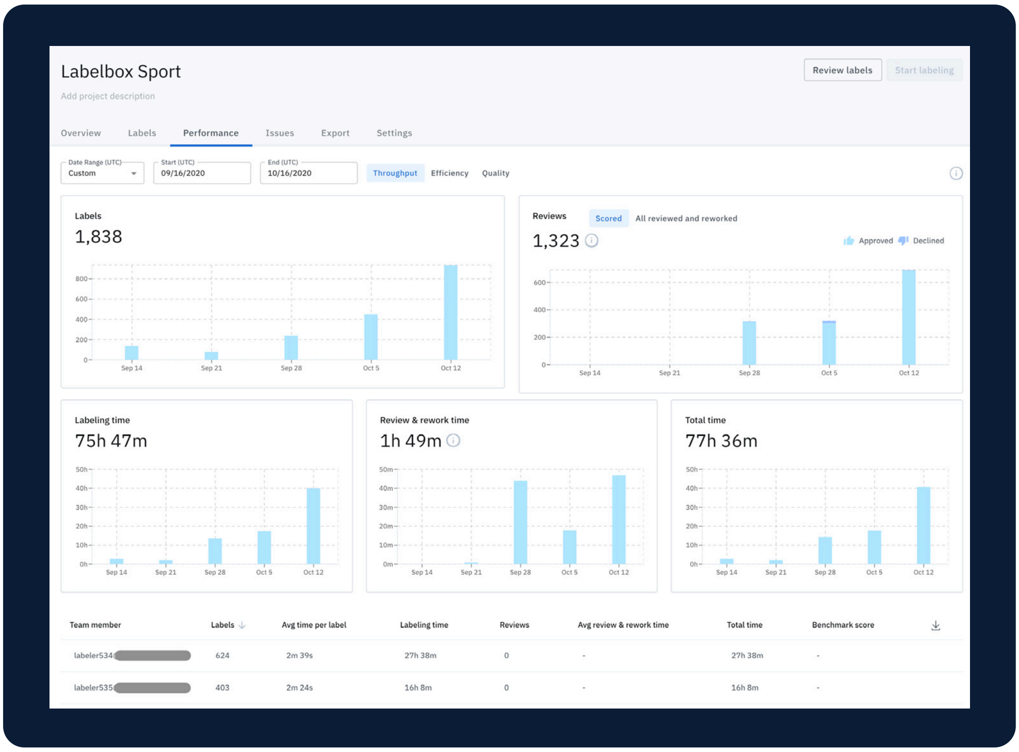
08 Monitoring labeling performance across project levels and individual levels

We've learned that organizations that routinely monitor and improve their labeling operations save significant time and costs and are better positioned to not only reach production AI but scale their AI efforts. There are three categories of metrics (inspired by those used by real-world manufacturers) to optimize labeling operations: throughput, efficiency, and quality.

Throughput is the total amount of training data produced in a certain time frame. Throughput metrics include:

- Per labeler throughput which monitors low and high-performing labelers
- Time spent labeling and reviewing, which keeps account of labeling costs and can be used to predict future project costs

Efficiency is calculated by comparing the throughput to the time/cost it takes to create training data. ML teams can either average all the labelers and gauge project level efficiency or track individual labelers. Tracking individual metrics can be particularly useful to determine whether the fastest labelers are also creating accurate labels, as well as to identify low-performing labelers for more training and guidance.

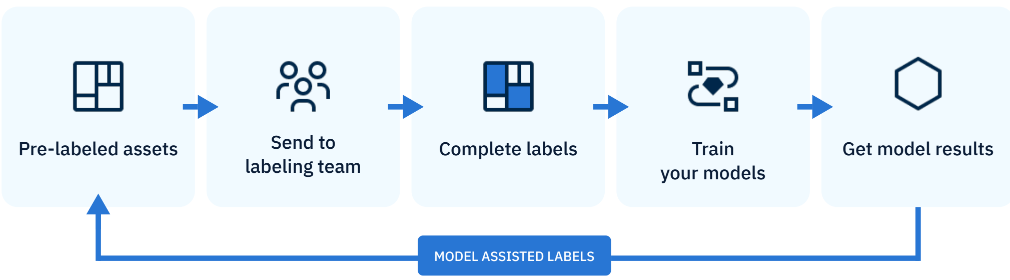


The performance dashboard showing throughput metrics in Labelbox.

And while we've discussed at length various ways to orient around quality for training data, ML teams should focus on the consistency of annotations as well as accuracy, or how labelers' annotations compare to ground truth, or benchmark assets.

09 Utilizing model-assisted labeling

Another essential practice used by leading AI teams is to incorporate automation for greater labeling efficiency. One way to achieve this is to leverage your own machine learning models through model-assisted labeling (MAL).

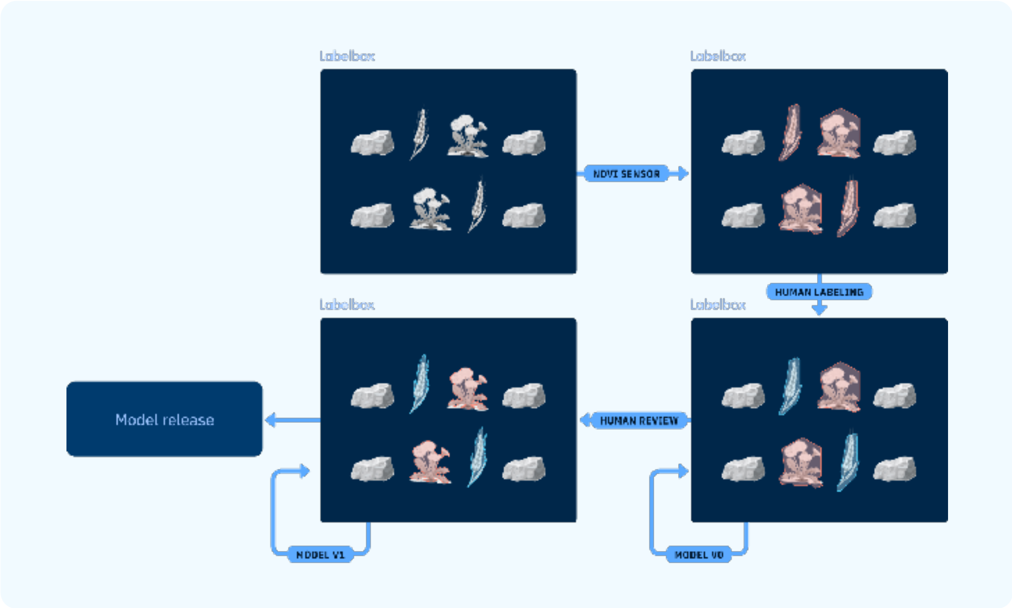


The model-assisted labeling process.

By using your own model to pre-label data, labeling teams can simply review, accept, decline, or refine model predictions rather than spend the bulk of their time creating annotations from scratch. This more efficient, faster workflow reduces the strain on labelers and opens up their bandwidth to make more impactful decisions. With each iteration, the amount of tedious work required by humans decreases even as the amount of training data increases. So, labeling teams can devote more time to low-confidence predictions and improving model performance.

Consequently, with fewer manpower hours needed from domain experts to label data, teams utilizing model-assisted labeling can save on annotation costs.

Annotating the tens of thousands if not millions of assets necessary to achieve a comprehensive data set has blocked many promising AI initiatives from ever reaching production. Model-assisted labeling is one of the simplest ways to reduce the waste of precious capital. In some cases, we've seen ML teams save 50-70% on labeling costs by utilizing MAL.



This agtech company cut its labeling costs by 50%. They import their model predictions into Labelbox where their labelers and specialists refine and complete reviews, boosting model performance.

By spending more time on model performance, you can additionally employ your model to generate labels with more accuracy. Your model is learning to identify examples with increasing accuracy. Model-assisted labeling also provides you with an early indication of the model's weak points. You get a quicker realization of what instances are difficult for the model to understand and may mislabel. Having the chance to improve or work around these weak points early on can be very beneficial.

Labelbox customer story: Blue River Technology

Blue River Technology creates intelligent machinery to aid farmers. They empower their farmers to implement more sustainable solutions, limit their chemical usage, and reimagine their routine processes to improve their farming yields year after year.

“Artificial intelligence and machine learning are not replacing the workers, but rather making them more efficient. We’re able to see things that we couldn’t see before and utilize our tools to get the most out of every plant. Blue River has been using computer vision with machine learning to provide targeted plant care,” said Christian Howes, software engineer at Blue River Technology.

Their See & Spray technology allows the farmer to operate large sprayers at the same speed as before, but is able to see over 200 feet of land and decide where the spray is needed and where it is not. Now growers can save money and make farming more sustainable. Those who have worked in agriculture know that weeds can be very hard to identify. Weeds have evolved over time, in some cases to look very similar to the crop.

Blue River Technology has adjusted how they capture data and uses overlays of different image formats to assist their labelers in determining what they are seeing in the images.

Blue River uses the Labelbox model-assisted labeling features to attach predictions to their images. Since the introduction of the model-assisted labeling, Blue River Technology has seen an overall reduction in labeling times of approximately 50%.

“Drawing all the outlines between plants is pretty time-consuming. And so model-assisted labeling has been a huge win for us,” said Howes. “Blue River was founded on bringing AI and computer vision to agriculture. Without AI, Blue River would not be here. We are still striving for our goals to make agriculture more sustainable and productive.”

Watch the Blue River Technology Breakthrough Customer of the Year Video to hear more on how model-assisted labeling is powering up computer vision projects at Blue River.

10 Spotting new correlations and identifying trends in data

Leading AI teams are set up to solve new problems and unanticipated situations that inevitably come up. It’s no surprise that changes to real-world conditions will require updates to your ML models in order to keep them relevant and account for edge cases. For example, facial recognition models needed to be updated as more faces suddenly wore masks once the pandemic began. These instances can confuse models and often necessitates a change in approach. The most successful teams have been trained to spot new correlations and unexpected outcomes.

One way teams are unlocking meaning in data trends is by enriching their training data with more metadata. Metadata provides a wealth of potentially valuable and untapped data. High-performing ML teams are adept at analyzing and leveraging metadata layers to extract actionable information. For instance, ML teams can use metadata to:

- Enable associations between datasets
- Enable reproducible model training
- Compare performance differences between models
- Understand the origin of artifacts
- Trace data lineage
- Detect data drifts when the production data distribution changes over time

Another way teams can visually find patterns and identify edge cases in data is through the use of model embeddings. By clustering visually similar data, teams can better understand trends in model performance as well as data distribution. While teams can calculate and plot clusters manually, some use cases in ML require more time-sensitive trend detection and a quicker approach. This is where a visual embeddings tool can be helpful to boost model performance.

Model embeddings in Labelbox can help visualize trends in data and cluster similar data, helping you uncover outliers faster.

Another downside to static graphs is the limitation of how much one can really explore the data, investigate groupings, and identify connections between variables. Leading AI teams employ interactive embeddings tools in order to specifically overcome those limitations and to help teams get more out of their data.

11 Utilizing humans on edge cases

Small mistakes in edge case annotations can lead to faulty decisions in cases of real consequence, especially in the life-saving realms of cancer cell detection or pedestrian detection for autonomous vehicles. We've seen leading AI teams adopt effective edge case management with a human-in-the-loop (HITL) process that significantly helps prevent small errors in training data from turning into massive mistakes.

Finding the right balance between human intervention and machine autonomy can be difficult for most organizations, one that often means the difference in scaling AI efforts and stalling. Even with the most well-defined algorithms and rules, once your model is exposed to real-world challenges, it is presented with an unfolding string of edge cases that need to be handled efficiently and correctly. Having a HITL workflow can allow for quicker, more efficient operations since the model can learn incrementally to be more accurate from the hands-on intervention.

Many industries utilizing machine learning implement HITL processes, such as cybersecurity, financial institutions, healthcare, manufacturing, and many more. For example, Boeing and Scotland-based company Anomalous are looking to develop potentially life-saving technology to detect aircraft faults quicker and more efficiently than relying on the human eye alone.[3] These life-threatening mechanical errors go undetected during manual inspections. By utilizing AI-powered cameras trained on how parts should look under varying conditions, the cameras can scan the aircraft's engine for defects quickly. A human in the loop then reviews the flagged information and determines whether or not the part is safe.

In addition to preventing detrimental errors and speeding up manual processes, having a HITL process in edge case management adds an important layer of transparency to your ML processes. As the general public becomes more cognizant about how enterprises are utilizing AI and ML, skepticism will invariably rise. ML teams will need to be able to take ownership of their models and HITL adds a layer of confidence both internally, and externally.

Enabling human annotators to take quick action on edge cases is one of the key advantages of using a training data platform for both annotation and data management. Under a single platform and interface, low-confidence predictions can be automatically prioritized, queued, and assigned to a domain expert to be solved given a specific time period.

Labelbox customer story: Genentech

Genentech is a biotechnology enterprise that has developed breakthroughs in medical research and medicine since 1976. Today, Genentech researchers are building convolutional neural networks to help diagnose illnesses and aid medical professionals.

While traditional, classic algorithms can perform well with "perfect" data, actual patient data is often riddled with small abnormalities.

[3] *Software expert takes on the future of aircraft safety by Hamish Burns, 2020*

"For the retina, there are about 10 different anatomical layers. We really need to delineate those layers for the classical method. We have solved that for healthy patients like a normal eye with a normal retina because that's a very easy problem and there's a lot of different classic algorithms. You can make it work," said Miao Zhang, Senior AI Scientist at Genentech. "But when it comes to real patients, there are a lot of different lesions and unexpected things. Basically, whenever you see a patient, the classical method will fail."

Training deep learning algorithms to find and classify images taken from real patients often requires meticulously labeled medical imagery numbering in the hundreds and even thousands.

Typically, only trained medical experts are trusted to correctly annotate training data for these use cases, as inaccurate model predictions can result in misdiagnosis and even loss of life. Labeling the required amount of data this way, however, is a costly and time-consuming endeavor.

"Deep learning has really started to change this field. Convolution neural networks can really solve the segmentation problem of patients. I would say any disease that needs imaging as a clinical endpoint or at least as a system seeing the diagnosis will see a big change from AI," said Zhang.

Genentech has revolutionized its ML process by having domain experts train teams of labelers on medical imagery annotation tasks at rapid speed. With labelers creating annotations, which are then sampled and reviewed for quality by experts, Genentech has developed a much faster and less expensive way to create training data and get their life-saving algorithms to production.

Watch the Genentech Breakthrough Customer of the Year Video to learn more about their story.

12 Putting it all together

We hope you've been able to come away with a few tangible ideas on how to better create and manage training data for your next ambitious AI project. To streamline these best practices, we've seen an increasing number of leading AI teams incorporate these training data practices into their workflow through the use of a training data platform (TDP).

A TDP is engineered specifically to help you improve your training data iteration loop, allowing you to label data quickly and accurately with a configurable workflow and automated labeling.

If you'd like to learn more, you can learn more about how Labelbox can accelerate your ML team's path to production AI by requesting a free demo. Our team will be happy to show you how our platform works and discuss how it can best be applied to your use case.

You can also try out our platform for free and get started immediately. Get started here.