Labelbox

Building labeling operations into your core MLOps strategy

00   Content

01   Overview

Machine learning teams often require large volumes of labeled training data in order to build production-grade AI applications. Building a labeling operations process that can deliver quality and quantity can be an arduous and frequently underestimated challenge, but it's an essential part of delivering the required training data. Having worked with hundreds of leading ML teams in this domain, we've found that it often takes months, if not years, to build these capabilities. This speaks to the importance of adopting best practices on how to incorporate this into your greater ML operations (MLOps) strategy from the onset. In this whitepaper, we will cover the key strategies to optimizing your labeling operations to serve your ML team with accuracy and efficiency. In addition, we will share how to more effectively evaluate and work with data labeling workforces as part of your overall ML operations.

02   Developing your labeling operations team

Enterprise ML models are usually built to solve specific, complex problems, so your labeling operations team should include domain experts who understand these problems and the issues you might face when addressing them. Labeling operations leaders should prioritize the voices of these subject matter experts when making decisions about the overall process.

If you're looking to build an in-house labeling operations team, begin by looking for people with experience in managing data labeling projects. While finding someone who has this exact experience may be difficult, you may find people who have been working as annotators in roles such as team leads, reviewers, or QA for labeling tasks. These types of roles are beneficial because they will understand the bigger picture of data labeling, such as which workflow to use, what processes to put in place, and how to adapt training when it comes to specific use cases or different data type formats such as video, image, or text.

Alternatively, if you choose to outsource any part of your data labeling, having someone with vendor management skills will be an asset. When outsourcing your labeling to a partner, you'll need to find the right one depending on your use case, and then you'll need to negotiate a contract with pricing and create a trusted relationship. Outsourced teams are typically located abroad so you'll need to account for time zone differences, pricing negotiations, quality controls, and regulatory compliance.

A key part of setting up your labeling pipelines and workflows comes in the form of project (or program) management. Each labeling task is a project in itself with a timeline for delivery, a budget to define, a quality target, and internal and external stakeholders who need to be

updated on progress. Specifically, you'll want to provide a detailed and thorough project set up from the beginning, saving time and money throughout the project's lifespan.

Additional things to look for in your labeling operations team include:

1. Which communication channels to use
2. Whether or not your team has a good understanding of technical requirements
3. If your team's quality control measures are focused from the start

Below is a detailed list of what a dedicated labeling operations project manager can help you with. Depending on your organization, this can either be in the form of a specific person(s) on your internal team or specialized team that is contracted with this experience:

- Create/optimize the data labeling pipeline for your data type
- Create/improve ontologies and guidelines for every labeling project
- Connect each project to the most suitable labeling team and set up communication channels
- Create/manage training processes to reach high quality labels
- Monitor quality while in production and flag/resolve potential quality issues
- Monitor and maintain a reasonable average processing time per label
- Reduce/impact overall labeling cost wherever possible

03   Building a portfolio of trusted vendors

The first objective to finding trusted vendors is to get a sense of which data types your team will label (e.g., video, image, text, audio, geospatial data) given that each instance requires a slightly different pipeline and workflow. You'll see that every labeling vendor has their own strengths and areas of speciality, so it makes sense to inquire and evaluate the past work that they've completed.

Image/Video	Text/Audio	Industry knowledge	Specialists	Certifications
• Satellite/Aerial • Segmentation • Bounding boxes • Classification	• Sentiment analysis • Content moderation • Named entity recognition • Classification • Translation • Transcription	• Insurance • AgriTech • Retail • FinTech • Customer support • Healthcare • Defense	• Foreign language experts • US-based security clearance • Medical doctors • Biologists • Chemists	• Soc 2 Type II • GDPR • HIPAA

Every vendor has their own strengths.
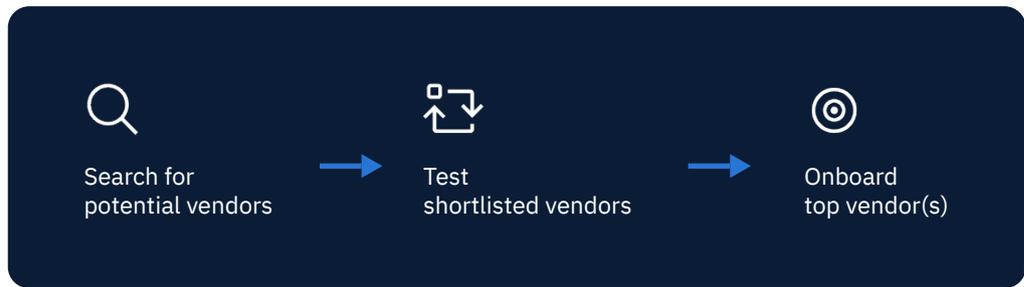
Vendor search & onboarding process

Labeling providers today cover a surprisingly wide spectrum of subject matter expertise. With some research, it is possible to find one that offers a specialized annotation service capable of labeling your dataset at a fraction of the cost it would take to hire and maintain an internal team.

Below are the steps required to identify and evaluate labeling vendors:

1. Search for potential vendors
   - First source via internet search, professional networks, inbound, etc.
   - Set up initial calls and ask the right questions to vet your vendors. Here are some topics you'll want to consider:
     - Full-time versus crowd
     - Flexibility/availability
     - Skills & expertise
     - Pricing
     - Communication style
     - Compliance needs and certifications such as SOC 2 Type II, GDPR, or HIPAA, along with any ethical commitments
     - Time it takes to kick off a new project

2. Initial test of shortlisted vendors
   - Set up a small labeling project with test data
   - Work with the vendors as you would in production
   - Check for quality, speed, communication, and project management

3. Onboarding your top vendor(s)
   - Put requirements in writing
   - Provide access to tools and training
   - Keep the vendor updated on upcoming projects

04   Defining your labeling process

After you have a shortlist of vendors identified, you'll want to start with a skeleton for what your labeling process will look like. Keep in mind that the labeling process is highly iterative; it is very rare that you will get high-quality labels from day one. You should expect there to be feedback loops, rework, and corrections several times before you hit a high accuracy rate.

This process typically includes:

1. Project ontology & data
   • Data:
     ◦ Selection
     ◦ Pre-processing
   • Ontology:
     ◦ Labeling efficiency
     ◦ Minimizing risks of human error
2. Labeling instructions
   • Written guidelines as reference for labelers throughout the process
   • Includes step-by-step instructions, labeled data examples, common mistakes
   • Aids in consistency across the labeling team
3. Preferred communication channel
   • Fosters two-way communication (Q&A, feedback)
   • Can be a document, group chat, regular meeting, etc.

Proccess flow of a new labeling project

05   Tracking metrics for quality, speed, and cost

Getting a solid understanding of the efficiency and quality of the training data being created is the first step to successfully getting the output needed. Let's cover the metrics that matter most: quality, speed, and cost.

Quality	**+**	**Speed**	**=**	**Cost**
• Vendor selection phase • Training process per project • Feedback loops / iterations • Quality control • Eyeballing		• Set up your timeline for delivery • Set up an expected time per task • Create milestones • Time difference between training and production phase • Track time per task, labeler, and project overtime		• Assess quality on an ongoing basis • Track time on an ongoing basis

Quality

Labeling at scale without compromising data quality requires transparency and visibility throughout your entire labeling pipeline. Adapting the training process to each new labeling project will also make a big difference because training your labelers will differ based on the complexity of the task and the desired output. Given the iterative nature of the labeling process, we suggest building in a series of feedback loops to allow for rework and corrections when aiming for a high accuracy rate. Creating best practices when it comes to quality control will help catch errors early on, triggering retraining if needed or changes to your guidelines in order to cover any grey areas or add new rules.

A few quality processes that you can put in place include:

- **Consensus and/or benchmark**
  Consensus features can help add several labelers to the same task to assess consistency among your team of labelers. This can be used to get votes on a subjective task or create multiple instances of the same asset. Benchmark features will help you compare your ground-truth data against labeler annotations. This creates golden standard assets that allow you to assess accuracy among your team of labelers.
- **Thumbs up or down**
  One additional way to ensure quality is to have ways for labelers to give feedback. For instance, the ability to "thumbs-up" the labels that meet your requirements and "thumbs

down" the ones that are incorrect provides a valuable feedback loop that is effective, systematic, and simple.

- **Issues and comments**
  Having the ability to ask questions and flag problems within the labeling workflow, while sharing well-labeled examples with the labeling team, is important for quality.
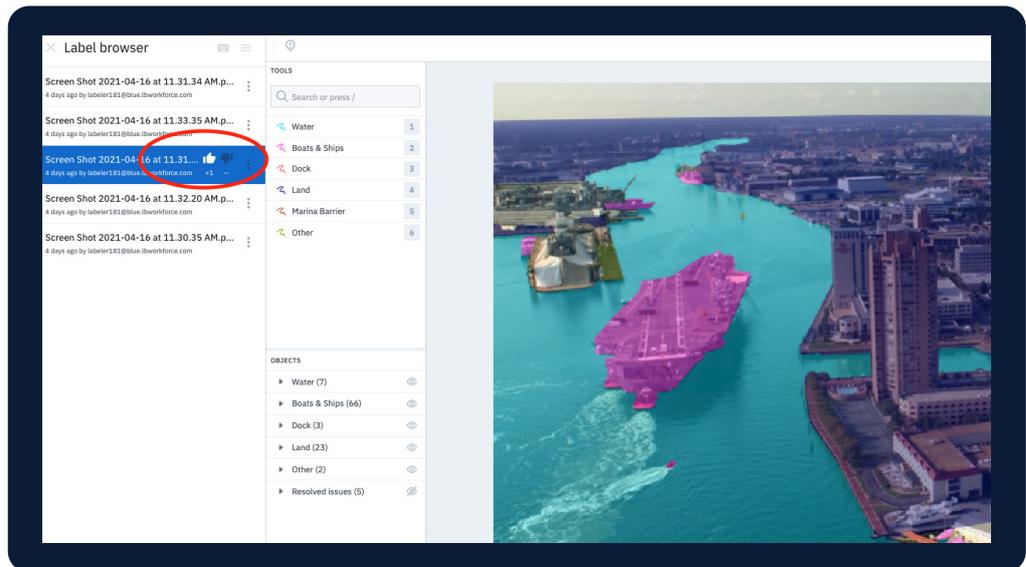
Labelbox enables reviewers to quickly and easily give feedback on annotations. They can give each label a thumbs-up or thumbs-down, and if further explanation is necessary, they can also leave comments.
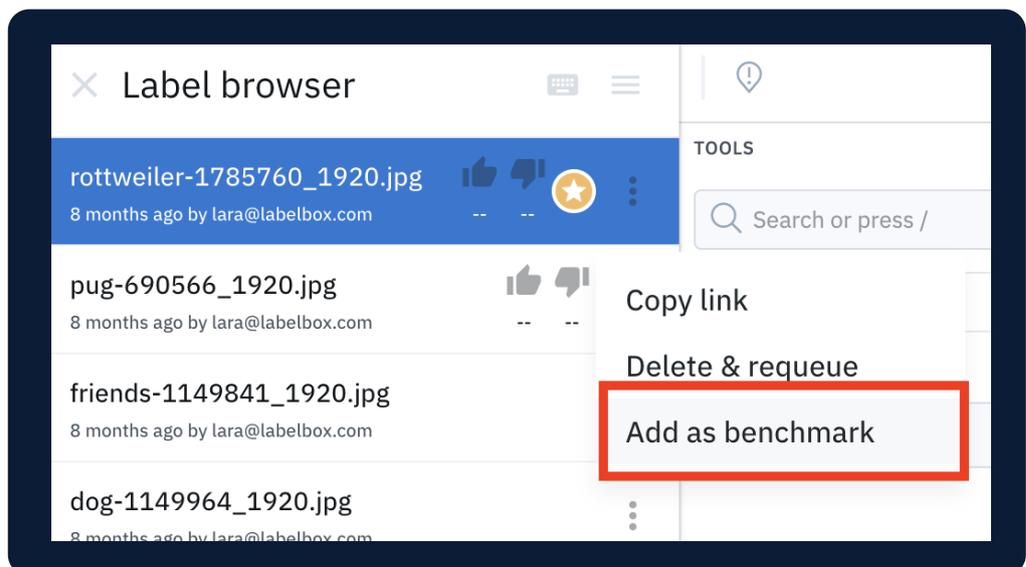
Labelbox allows teams to mark ground truth labels as benchmarks. Other labeled assets can then be compared to this benchmark to determine whether the label is accurate.

Speed

Speed is a useful metric to track as this will have a direct impact on your overall cost. Defining a timeline for delivery with your labeling partner will help avoid misunderstandings and give your vendor a deadline to work towards. A deadline based on a reasonable average time per task or asset will help your labeling partner reach your requirements early on and create a baseline of mutual expectations. When assessing labeling efficiency, be sure to specifically look for anomalies and trends.

We recommend evaluating speed at the project level, the task level, and the labeler level over time. This will provide enough information to identify any red flags. For example, if the timing is consistently missing the mark, you might need to review your forecast and inform your stakeholders.

Additionally, you might want to look into the annotations of a very slow or very fast labeler to check the quality of their work. Timing outside the norm is a great key indicator that a project, task, or labeler requires some additional attention.

Read on for a few recommended practices that will help your team track speed.

Project dashboard
As a best practice, we recommend setting up a dashboard to track the health of your project. This includes tracking time per task, time per labeler, and the volume annotated every day, which will help you get to a more accurate forecast for completion and throughput.
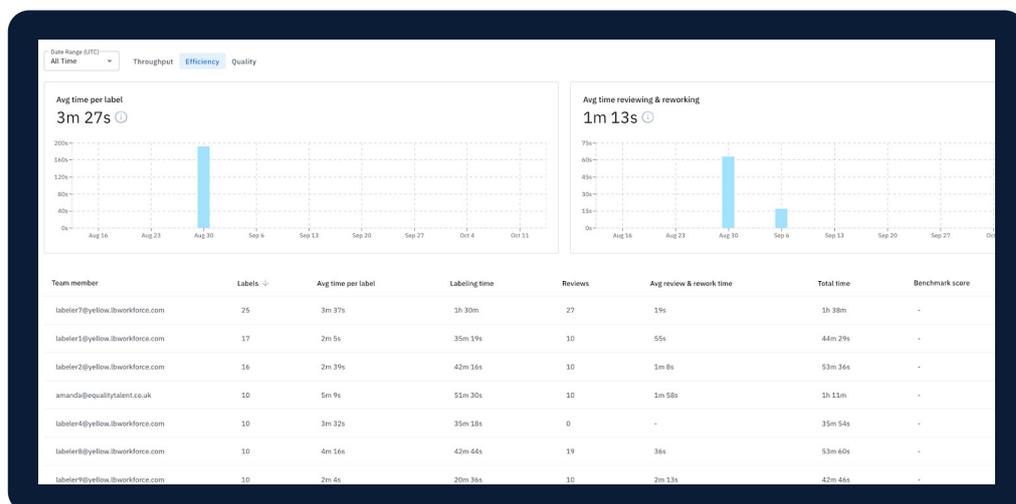

Labelbox enables teams to track labeler performance with metrics, including average time per label.

Time tracker
The average time per label (also known as time per task or TPT) measures how long it takes an individual to complete the initial label. This is a direct measurement of labeling efficiency. Look to see if there are any variances in the average time per label and review for upward or

downward trends. We've found that downward trends are typical for a labeling operation, as labeling efficiency often improves 30% or more over the first six weeks. The "average time reviewing and reworking" metric measures how long it takes an individual to perform a review and/or rework (edit) the label.
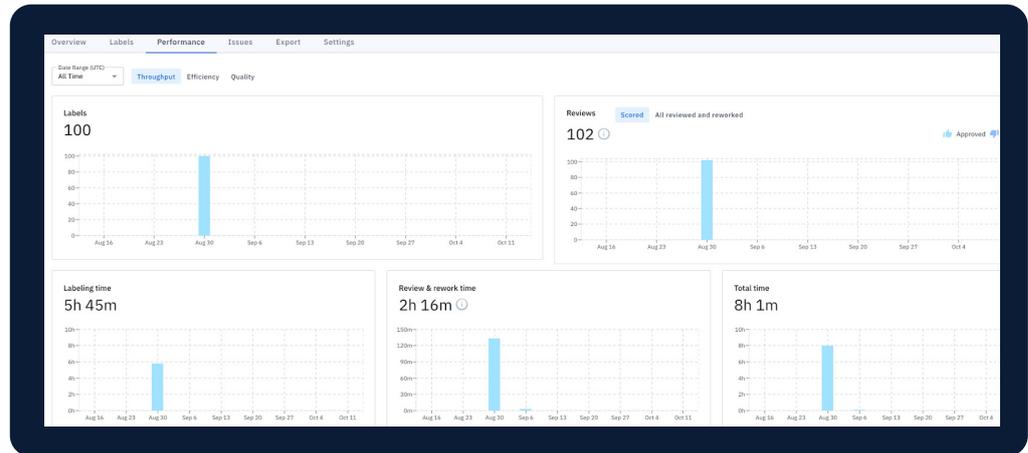
Teams can track throughput metrics such as labeling time, review and rework time, number of labels completed, and more.

Cost

All of your top line metrics will roll into your overall costs. Cost will be impacted positively if you focus on producing high-quality annotations at a reasonable average processing time per task. Keeping track of these metrics over time will help you estimate your costs and timelines to make sure your projects are staying on deadline and on budget.

Overall usage tracker

We suggest implementing an overall usage tracker, which will help you get real-time insights into time spent across all your projects and allow you to control your spending.

A usage tracker enables an ML team to get real-time insights on time and costs across all their projects.

06 How to scale up the entire labeling pipeline

Helping your workforce scale up

First, we recommend running a forecast report to assess if hiring more labelers is necessary to scaling up your labeling. If it does require hiring, we have found that two weeks is a reasonable amount of time to hire and expand a team for your production project and to get to the right size team.

1. Interview and hiring process
2. Initial training (on the labeling platform)
3. Specific training on your project
4. Testing on your project

How a typical interaction between a program manager, labeling trainer, and QA team will look.

Maintaining quality while scaling

It's important to evaluate your workforce based on their ability to scale up or down based on your organizational needs. In most cases, you may have to label data in real time within a timeframe, so it's important to discuss these capabilities with your target vendors beforehand. The service level agreement (or SLA) you choose will depend on the importance of timely data in your production processes as well as quality expectations.

In addition, prior experience working across specialized labeling tasks is usually an advantage in terms of the time it will take to ramp up the outsourced labeling team. Ask whether the provider can offer industry experts if your project requires specialized expertise.

Key events such as new product launches can generate spikes in data labeling volume. Depending on your AI application, seasonality may also play a large part in volume as well. It's important to discuss capabilities with your vendor and understand what the ramp-up times look like for adjusting your labeling needs. We recommend understanding whether workforces can accommodate for any spikes of work that may arise. In addition, be sure to increase your workforce gradually over time to mitigate some of the learning curve costs that you'll incur, and ensure that you or your labeling service provider monitor per-labeler metrics such as quality and speed as you scale.

We recommend the following stages when integrating new labelers to an existing team:

1. **Training:** Set up a project dedicated to training
2. **Testing:** Set up another project for testing
3. **Performance management:** Quality checks are performed during production by a team of reviewers that will assess the entire team of labelers
4. **Additional QA:** Give more attention to labelers that are new to the production team

07 Expertise made effortless

Labelbox enables ML teams with the tools to track throughput, efficiency, and quality metrics, integrate quality management workflows into your labeling operations with ease, scale your team and datasets, and much more. If you and your team are strapped for time or resources and don't want to manage all of this complexity internally, Labelbox Boost can help. With

Labelbox Boost, we can handle data annotation for you, from finding the right workforce teams to managing the entire process and the teams themselves.

An extensive array of Labelbox Boost customers today rely on the service to drive their entire labeling process. In addition, the team can help you keep costs low by driving efficiency during task design and labeling while offering the flexibility to ramp the labeling team up and down with your data volumes.

Our Boost team has helped hundreds of customers develop their labeling operations over time, both in partnership with the Labelbox Workforce and with their own labeling teams. Today, you can engage Boost in two ways:

1. **Workforce Boost:** Our experienced team will find the right labeling workforce for your needs and help you manage them throughout labeling projects.
2. **Complete Boost:** Labelbox becomes an extension of your team as we program manage your entire labeling workflow across multiple projects (even as your scope or needs evolve), including ontology creation, labeling instruction setup and improvements, quality monitoring, and more.

Labelbox was built to make the labeling operations process easier and to empower enterprise machine learning teams to speed up their model iteration and time to market. With Boost, leading AI teams increase productivity, keep costs low, and the control and flexibility of dedicated project management and instruction design without building it from scratch. To learn how Boost has benefited other ML teams, check out our case studies with Move.ai, American University of Beirut, and the BioMedIA research group.

Labelbox

Learn more about our offerings, sign up for a demo, or start using our free version today at www.labelbox.com.