

Choosing the Right Data-Labeling Solution

One obstacle data scientists, developers, analysts, and line-of-business decision makers face at the start of building and operationalizing natural language processing (NLP), and realizing the value of their investment, occurs with data labeling. Our customers consistently tell us that labeling documents is a time-consuming, repetitive, and error-prone task and they need a better solution.

Better labeling means better data on which to train NLP models, and therefore higher model performance and faster deployment of models into production. One hundred great labels can mean the difference between a model transforming your business, or it just being another underperforming prototype.

Building good NLP models that you can trust, especially for our mission-driven customers in national security, intelligence, and the Fortune 1000 customers is vitally important.

Data labeling designed for your NLP workflow

Encoding human domain expertise into an NLP model makes one person's domain expertise infinitely scalable. But domain experts are expensive and have more valuable things to do with their time than label data. Our customers want to minimize the time spent labeling data while maximizing model performance.

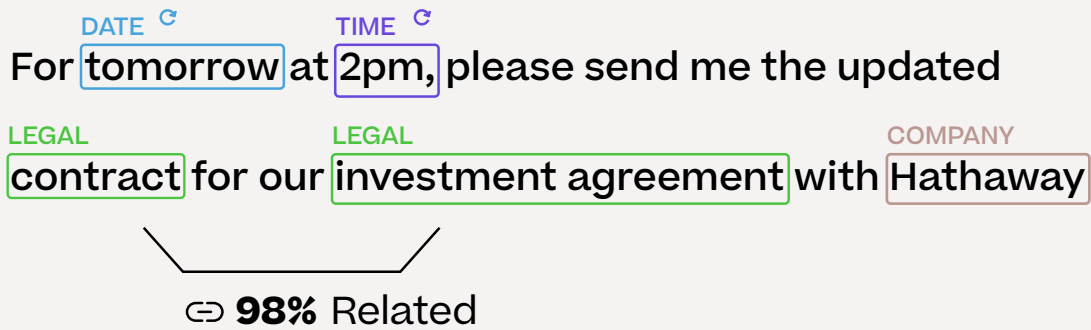
Primer offers three data labeling solutions that dramatically shorten the labeling process while increasing the quality of your dataset:

- Team-based labeling using LightTag
- Automated labeling using LightTag and Primer Automate
- Outsourced data label through our partnership with Odetta

1. LightTag: Team-based data labeling

Primer offers LightTag to manage, expedite, and ensure the quality of your labeling. It is available as a standalone product, SaaS, or on-premises for our security conscious customers. It incorporates single sign on, role-based access controls, and audit logging to support compliance.

LightTag's team-based labeling solution offers innovative features for project management, annotation scoring, conflict management, and quality control. LightTag makes it easy to set up labeling tasks for teams of labelers, accelerating the time it takes to start training NLP models. This can be done online, with labeling starting in minutes. If you are training a model on sensitive data, or need to keep your data in house and private, you can deploy LightTag on premises in a secure environment.



Control Your Data Quality

LightTag's Review Mode and Reporting make it easy to ensure your data is labeled accurately and your annotators are performing at their best.

LightTag can visually indicate conflicts between annotators and has the ability to automatically accept annotations by unanimous or majority vote. The speed and scalability of LightTag can result in a model in less than 24 hours, greatly reducing time to value.

LightTag's UI is geared toward project management, saving time and effort for those who currently use Excel or Google sheets to manage numerous cross-team projects. Instead of cleaning and labeling data sets, data scientists are free to focus on the key problems they need to solve.

Once annotation is completed and the model is deployed, you can monitor its performance on production data in LightTag. By quickly validating model predictions, you can both monitor performance, and get new labels to continuously train and improve the model. LightTag's integration with Primer allows you to validate models during development and as they are deployed, to ensure that predictions are correct and compliant, and facilitate retraining to optimize performance.

2. LightTag + Primer Automate: Automated data-labeling

Get the flexibility and accurately labeled data you need to train your NLP model. LightTag and Automate, a no-code, end-to-end artificial intelligence solution for NLP model creation and deployment, are integrated. Automate's intuitive labeling interface makes labeling and other NLP tasks easy. With Automate, you can go straight from labeling to deployment with an all-in-one, easy-to-use system. You label data, Automate builds the model and deploys it, saving headcount and time on model development and production. Configure, label, train, and deploy your models and get started using AI right away — no coding or technical skills required.

Both LightTag + Primer Automate and LightTag team-based labeling offer superior management capabilities, optimized to work with a pool of known annotators that can be efficiently managed, as opposed to labor pools of unknown background. This difference improves expressivity, management efficiency, and quality control.

3. Odetta: Outsourced data-labeling

Some customers don't have the time or resources to label their own data. Outsourcing saves time and money, yet you want to make sure the annotators are trusted, reliable, trained, competent, respond to feedback, and feel like an extension of your in-house team. We offer a managed team of annotators through our partnership with Odetta.

Primer is a founding customer of Odetta; we've worked with them for years. With Odetta, we offer an ethical, fast, and accurate solution that accelerates your machine learning lifecycle. We have tried other labeling teams in the past, and we were not able to get the high quality we need to train top-tier models using the data produced by these labelers.

Odetta is an on-demand workforce whose services include data collection, data labeling, and data analytics. Their team consists exclusively of women, often college-educated in STEM fields, who are primarily based in South Asia and the Middle East. We partner with them because they provide high-quality data and reduce operational overhead by fully training and managing labeler resources.





Data labeling accuracy and its impact on training data quality becomes an evaluation of a data set's fitness to serve its purpose in a given NLP use case. To improve the quality of data labeling and text annotation, it must be measured accurately: Expressivity, Efficiency, Management, Quality control, Ease of integration.

Expressivity— Can you express the kinds of annotation that you need? We can express any kind of annotation, from document classification to complex relationships across texts.

Efficiency— Can you express the annotations that you need at an acceptable pace? You need good, labeled data fast. LightTag leverages UX best practices and AI learning to improve and ensure labeled data is delivered on time and on budget.

Management— Can you manage the annotation project as it scales in annotators, annotations, and complexity of annotation? As annotation projects scale, they become increasingly complex and need tailored project management solutions. Our workforce management system automates project management overhead giving you back precious hours in the day.

Quality Control— Can you control the quality of annotations? Having labeled data is not enough, it needs to be exceptionally accurate and consistent. LightTag's quality control and review functionality enables rapid detection, prevention, and mitigation of data errors before they propagate to downstream processes.

Better data labeling, more trustworthy results

By combining Primer Automate's machine-driven programmatic labeling capabilities with LightTag's innovative human-driven labeling, and a full-service outsourced, white glove labeling solution, we offer the starting block of an end-to-end NLP solution that is robust, flexible, and adaptable to your organization's needs. Gaining strength at the starting line with labeling also means that with the benefits of better model training and model performance, faster model deployment, you can realize, and trust, the results of applied NLP results faster than ever.

Learn more

Want to learn more about our data labeling solutions, product demos, and free trials? [Contact us](#). Try [LightTag for free](#) and/or [send us a note](#) to discuss your specific data labeling and NLP needs.