



# Active Learning

What It Is & How Alegion Does It

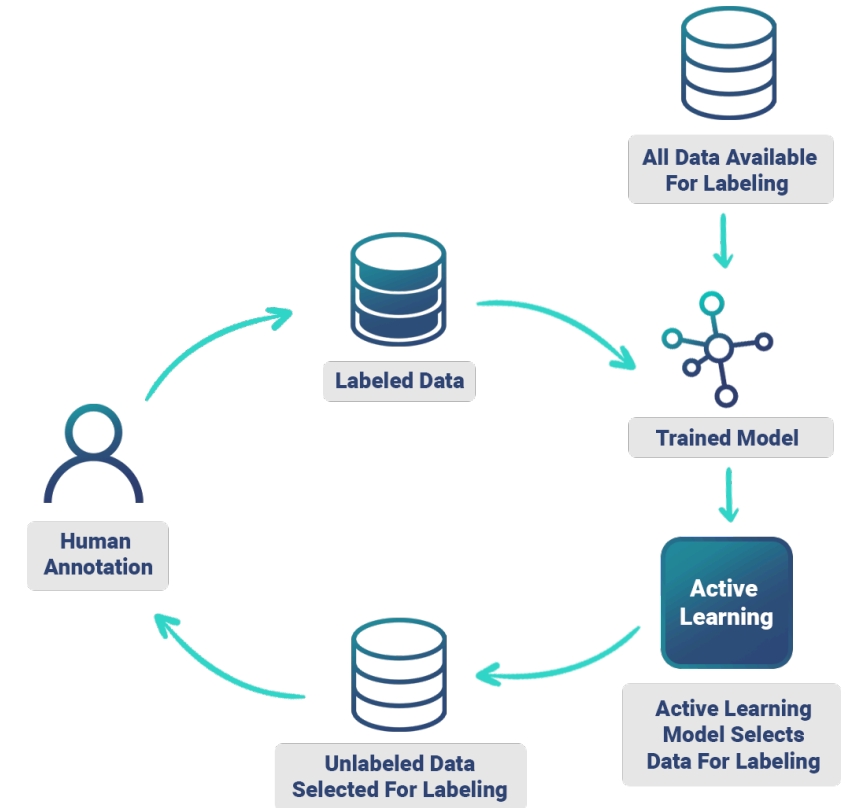


# What Is Active Learning?

Active learning is a methodology used in machine learning that helps focus data labeling on instances in your data set that will drive the greatest value for your model. Through a variety of algorithms and processes, models are able to identify subsets of valuable data, refer these subsets to human annotators, trigger model retraining with the newly labeled information, and drive greater machine learning model accuracy with less human labeling overall.

Active learning methods provide a more cost-effective path to a high performing model. Not every image or instance is worth the time or money it takes to label, but some instances are incredibly valuable once labeled well.

Active learning can be a useful tool at many stages of model development, and it is a powerful example of how models, sampling mechanisms, and human annotators can work together quickly and intelligently to produce quality results.



# How Active Learning Works

Let's look at video data from soccer games to explore how active learning can help drive a model forward.

Assume your model has already been trained using a small, labeled subset of your data. It can easily and with a high degree of confidence track the movement of the ball, classify hands and feet, ID individual players when they are standing apart from each other against the green background of the field, etc. However, your outputs show a lot of uncertainty (low confidence) when your model tries to classify and track movements of players' shoulders and hips or hold on to the appropriate ID for each player after several players have collided around a ball.

If active learning is incorporated at this stage of your model development, you can use the model's confidence levels to suppress all labeling of instances that your model already handles well and only work on annotation of complex or edge cases. Human annotators will annotate a lot of shoulder joints until the model has mastered those as well. Human annotations will help the model learn how to hold onto unique player IDs at each moment of a crowded scene and beyond.

Meanwhile, your model is moving quickly toward being production ready, and you are saving time and money by avoiding unnecessary annotation of balls, hands, and feet.



*This example demonstrates a common query method: uncertainty sampling based on model confidence. The result is better inferences on previously rare classes, but other methods can maximize model change, minimize generalization error, and quickly identify novel classes. These can be used individually or in concert. Consider also "query by committee", where sampling is based on the level of disagreement from multiple models.*

# What Are The Advantages of Active Learning?

There are three major advantages to using active learning as part of your model development.



## Particularly useful for large datasets.

Some computer vision teams are drowning in data and need help turning their huge, unlabeled datasets into something a model can begin to learn from. Large datasets can be very expensive to label and typically have lots of redundancy which makes the use of active learning methods all the more valuable.

No matter how large the dataset, the active learning approach selects or curates the assets that are most informative for your model and only passes these highly valuable assets on to human annotators.



## Drives accuracy and performance for special cases.

Your model has to be able to handle the unusual. Active learning loops help models focus on classifying edge cases and nuanced or complex examples, giving you the ability to see the large dataset and yet focus on the special cases. It can take your model from basic training to extremely high levels of confidence and accuracy for even the trickiest case.



## Saves time, money, and human effort.

This is the big one! The biggest advantage of active learning is that it gets your model to high degrees of accuracy, faster. Human annotators get to focus on labeling only highly valuable instances, and your model gets to focus on learning only what it still needs to know.



# How We Do Active Learning At Alegion

We use active learning to help customers with their model development when models need to get better and better at granular classification of low confidence instances.

When you work with Alegion to build or improve your model, inference confidence levels can be passed from your platform into ours via API, allowing easy automation of the process. With that data feed, our platform drives the model forward, selecting which assets to send for human annotation.

The Alegion platform and our data labeling services are designed to take your model from basic labeling to specialized, production ready labeling as quickly and accurately as possible. Our customer success team is with you every step of the way to provide expert guidance in how to get the most accurate data for the most accurate models by labeling the most important examples of a massive data set, and actively crafting and evaluating a multi-phase labeling effort to balance cost against quality, helping you save time and money.

*Curious what active learning could do for your model?  
If you have questions or want to learn more, please reach  
out to [solutions@alegion.com](mailto:solutions@alegion.com) and get in touch.*

