

Accelerating Machine Learning Model Development with Human in the Loop



AI promises to change our world, and fast. Recently-developed machine learning models can use the audio files from cell-phone recorded coughs to accurately detect coronavirus, even in people with no symptoms. Machine learning is creating unprecedented possibilities for innovation across traditional industries like healthcare, transportation, and agriculture.

However, developing ML models is a perilous journey. AI development teams encounter challenges including data scarcity, dirty data, workforce burdens, and algorithm and quality control (QC) failures. At each stage, they will make decisions about their workforce that can mitigate these challenges and, ultimately, may determine success or failure.

Developing models and maintaining them in production requires a lot of data and skilled people to work with it. It can be difficult to scale that process, and a cursory approach can result in high costs, low-quality data, and

poor model performance.

At CloudFactory, we've been processing data for over a decade. We've found the benefits of a human-in-the-loop machine learning approach begin with model development and extend across the AI lifecycle, from proof of concept to model in production.

In our latest whitepaper, we share how to strategically apply people to accelerate model development and the AI lifecycle. In this article, we share our perspective on human in the loop and accelerating the model development process.

What is human in the loop?

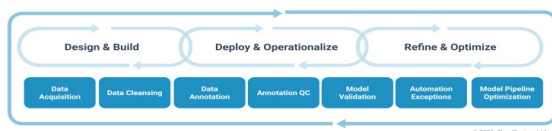
In machine learning, human in the loop (HITL) refers to work that requires a person to inspect, validate, or make changes in the training and deployment of a model into production. For example, the data scientist or data engineer who verifies a machine learning model's predictions before moving it to the next stage

of development is the human in the loop.

A similar approach extends to the people who collect, label, and conduct quality control (QC) on data for machine learning. About 80% of total project time is consumed with data preparation, labeling, and processing, according to analyst firm Cognilytica. The majority of that work is done by people.

HITL and the AI Model Development Process

There are seven key areas in machine learning model development where people who prepare, label, validate, and optimize data play key roles. Those roles are represented by the blue boxes in the image.



We break down the machine learning model development process into three stages: design and build, deploy and operationalize, and refine and optimize. The blue boxes show where people are involved. Accelerating that process requires a human-in-the-loop workforce that can adapt to changing tasks and volume.

Let's take a closer look at how you can apply people strategically at each of these phases to accelerate the model development process.

Design and Build

About 80% of AI project time is spent in this phase of the model development process, where the focus is on data acquisition, cleaning, and annotation. Here, a human-in-the-loop workforce can be used to acquire and prepare data for annotation.

Subjectivity and understanding of edge cases are important in nuanced work, such as data cleansing (e.g., reformatting, deduplication) and enrichment (e.g., web scraping, research). This process can be accelerated with a workforce that is able to adapt to evolving tasks and can add workers to increase throughput as volume changes.

Deploy and Operationalize

AI projects are put to the test in this stage of model development, where teams are deploying the model they've designed and optimizing processes for produc-

tion. In fact, about 96% of failures at this point in the process, according to Dimensional Research.

Accelerating the process at this stage may include automation, such as auto-labeling. It's wise to apply automation to proven processes. For best results, automation should be paired with a human-in-the-loop workforce to ensure it performs as expected and to manage exceptions and QC. Also during this phase, HITL workforces can validate the outputs of the model and assist with data tasks to determine model accuracy.

At this stage in model development, agriculture company Hummingbird Technologies outsourced their image annotation work and annotation QC to a professionally managed team so they could focus on supporting model improvements.

Refine and Optimize

In the final phase, teams begin thinking about what is needed for the model in production. Here, they are monitoring to detect and resolve model drift and ensure high performance. They also are thinking about how to handle automation exceptions.

Here, a workforce can accelerate the process by monitoring outcomes, helping to support a healthy data pipeline, and identifying and resolving exceptions. Of course, there likely will be more work to gather, annotate, and train the model again to respond to changes in the real-world environment, as we saw when weird behavior was breaking AI models when cities went into pandemic lockdowns, forcing humans to set them straight.

The Bottom Line

Most AI project time is spent on aggregating, cleaning, enriching, and labeling data. Training the model is undeniably a laborious task, which is why many companies developing AI models outsource these processes. Still, designing and building the algorithm is merely the first stage of the AI lifecycle. AI models require continuing oversight for quality control, validation, and optimization. By deploying a strategically managed workforce throughout every phase of the lifecycle, teams can scale and accelerate the process