A machine learning (ML) developer is an expert on using data to training models. The models are then used to automate processes like image classification, speech recognition, and market forecasting.

Definitions of machine learning roles can vary. Often there’s some conceptual overlap or even conflation with the roles of [data scientist](https://www.toptal.com/data-science/job-description) or [artificial intelligence (AI) engineer](https://www.toptal.com/artificial-intelligence" \l "hiring-guide). Machine learning is actually a subfield of AI that focuses on analyzing data to find relations between the input and the desired output.

A machine learning developer produces a tailor-made solution for each problem. The only way to achieve optimal results is to carefully process the data and select the best algorithm for the given context.

### Responsibilities

* Understanding business objectives and developing models that help to achieve them, along with metrics to track their progress
* Managing available resources such as hardware, data, and personnel so that deadlines are met
* Analyzing the ML algorithms that could be used to solve a given problem and ranking them by their success probability
* Exploring and visualizing data to gain an understanding of it, then identifying differences in data distribution that could affect performance when deploying the model in the real world
* Verifying data quality, and/or ensuring it via data cleaning
* Supervising the data acquisition process if more data is needed
* Finding available datasets online that could be used for training
* Defining validation strategies
* Defining the preprocessing or feature engineering to be done on a given dataset
* Defining data augmentation pipelines
* Training models and tuning their hyperparameters
* Analyzing the errors of the model and designing strategies to overcome them
* Deploying models to production

### Skills

* Proficiency with a deep learning framework such as TensorFlow or Keras
* Proficiency with Python and basic libraries for machine learning such as scikit-learn and pandas
* Expertise in visualizing and manipulating big datasets
* Proficiency with OpenCV
* Familiarity with Linux
* Ability to select hardware to run an ML model with the required latency