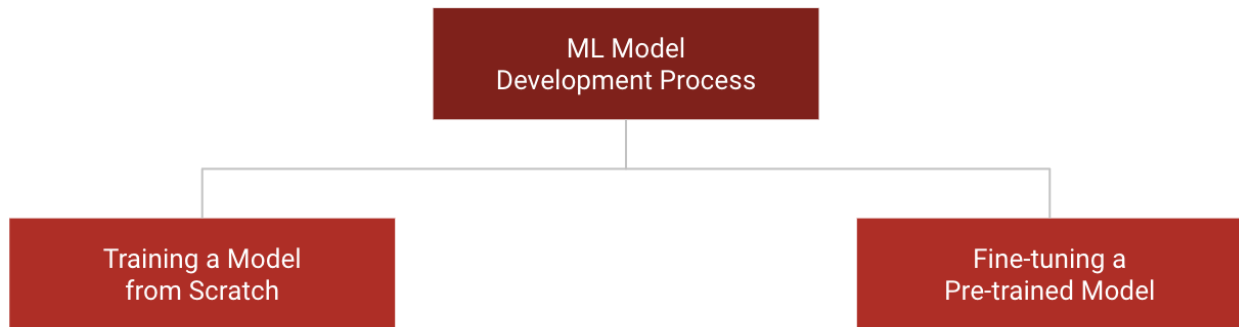# Training vs. Tuning



## About This Reading

Training a model from scratch might seem to be the norm. Certainly, this is almost always feasible if you have wealth access to a large corpus of data and training resources. But quite a few times, we are unable to train machine learning models from scratch. This might be due to the lack of computational resources, time available, or a paucity of available training data.

Ideally, we would like to minimize unnecessary computation to help save us time, money, and also to minimize carbon emissions as a result of our computation, but also to port existing models used for general tasks to our more specific application. To this end, this section will discuss the pros and cons of training a model from scratch vs. fine-tuning a pre-trained model.

## Examples

Imagine a scenario where you are developing an image-captioning model to produce a description of a patient's chest x-ray. You might only have access to a small amount of data to develop this, say, 1000 x-rays and reports, and thus training an image captioning model from scratch would produce poor results. However, we can take an existing, more general, image-captioning model and use this as the starting point for our model. Sometimes, this is called "pre-initializing" or "pre-training" our network, since we import the weights from a previously trained network. Following this, we can train the latter layers of our network (while keeping the early layers frozen, since these encode more primitive image properties) on our smaller dataset of chest x-ray data. The results from this implementation should far exceed those of a model trained from scratch on the chest x-ray data. Clearly, this only works when the two tasks are similar enough that the model can be ported from the more general application to the more specific application. To this end, repositories of models have been developed (e.g., Hugging Face) to store models that are freely available for use. Some examples of popular models are BERT, VGG16, Inception, and GPT-2. These models were developed by organizations with vast amounts of computational resources, and are thus prohibitively expensive for any individual to produce. However, the fact that we can use and adapt these models to our own applications is a triumph of open-source machine learning.