

PyTorch is an [open-source machine learning library](#) based on the [Torch](#) library,^{[4][5][6]} used for applications such as [computer vision](#), deep learning research^[7] and [natural language processing](#),^[7] originally developed by [Meta AI](#) and now part of the [Linux Foundation](#) umbrella.^{[8][9][10][11]} It is one of the most popular [deep learning](#) frameworks, alongside others such as [TensorFlow](#),^[12] offering [free and open-source software](#) released under the [modified BSD license](#). Although the [Python](#) interface is more polished and the primary focus of development, PyTorch also has a [C++](#) interface.^[13]

PyTorch utilises [tensors](#) as a intrinsic datatype, very similar to [NumPy](#). Model training is handled by an [automatic differentiation](#) system, Autograd, which constructs a [directed acyclic graph](#) of a forward pass of a model for a given input, for which automatic differentiation utilising the [chain rule](#), computes model-wide gradients.^[14] PyTorch is capable of transparent leveraging of [SIMD](#) units, such as [GPGPUs](#).

A number of commercial [deep learning](#) architectures are built on top of PyTorch, including [Tesla Autopilot](#),^[15] [Uber's Pyro](#),^[16] [Hugging Face's Transformers](#),^{[17][18]} and [Catalyst](#).^{[19][20]}

History

In 2001, Torch was written and released under a [GPL license](#). It was a machine-learning library written in C++, supporting methods including neural networks, [SVM](#), [hidden Markov models](#), etc.^{[21][22][23]} It was improved to Torch7 in 2012.^[24] Development on Torch ceased in 2018 and was subsumed by the PyTorch project.^[25]

Meta (formerly known as Facebook) operates both PyTorch and Convolutional Architecture for Fast Feature Embedding ([Caffe2](#)), but models defined by the two frameworks were mutually incompatible. The Open Neural Network Exchange ([ONNX](#)) project was created by Meta and [Microsoft](#) in September 2017 for converting models between frameworks. Caffe2 was merged into PyTorch at the end of March 2018.^[26] In September 2022, Meta announced that PyTorch would be governed by the independent PyTorch Foundation, a newly created subsidiary of the [Linux Foundation](#).^[27]

PyTorch 2.0 was released on 15 March 2023, introducing [TorchDynamo](#), a Python-level [compiler](#) that makes code run up to 2x faster, along with significant improvements in training and inference performance across major [cloud platforms](#).^{[28][29]}

PyTorch tensors

Main article: [Tensor \(machine learning\)](#)

PyTorch defines a class called Tensor (`torch.Tensor`) to store and operate on homogeneous multidimensional rectangular arrays of numbers. PyTorch Tensors are similar to [NumPy](#) Arrays, but can also be operated on a [CUDA](#)-capable [NVIDIA](#) GPU. PyTorch has also been developing support for other GPU platforms, for example, AMD's [ROCm](#)^[30] and Apple's [Metal Framework](#).^[31]

PyTorch supports various sub-types of Tensors.^[32]

Note that the term "tensor" here does not carry the same meaning as tensor in mathematics or physics. The meaning of the word in machine learning is only superficially related to its original meaning as a certain kind of object in [linear algebra](#). Tensors in PyTorch are simply multi-dimensional arrays.

PyTorch neural networks

Main article: [Neural network \(machine learning\)](#)

PyTorch defines a module called nn (`torch.nn`) to describe neural networks and to support training. This module offers a comprehensive collection of building blocks for neural networks, including various layers and activation functions, enabling the construction of complex models. Networks are built by inheriting from the `torch.nn` module and defining the sequence of operations in the `forward()` function.