Deep learning vs. machine learning in Azure Machine Learning

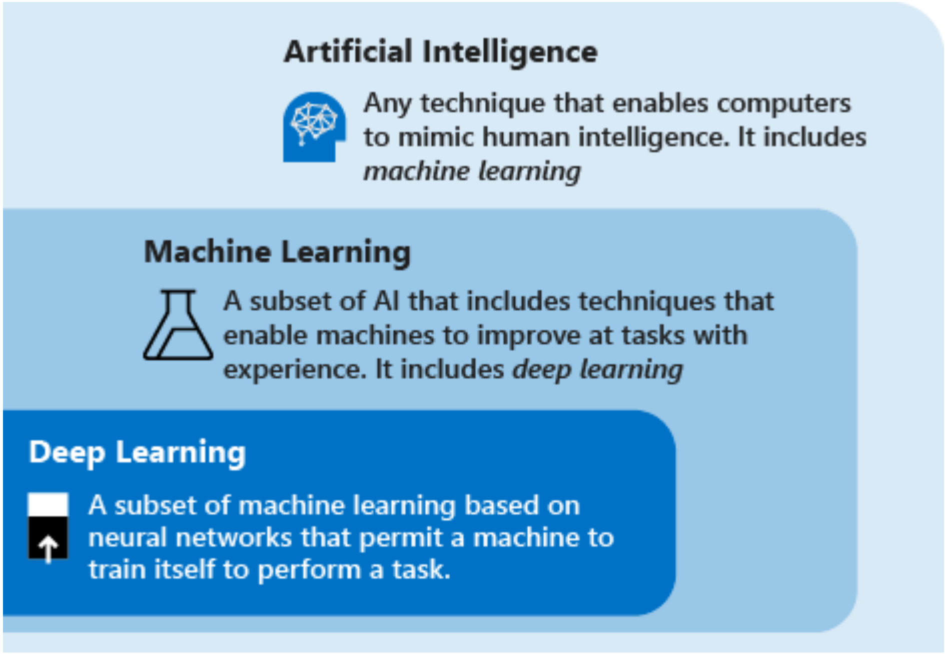
04/12/2021 • 9 minutes to read • +4

In this article

This article explains deep learning vs. machine learning and how they fit into the broader category of artificial intelligence. Learn about deep learning solutions you can build on Azure Machine Learning, such as fraud detection, voice and facial recognition, sentiment analysis, and time series forecasting.

For guidance on choosing algorithms for your solutions, see the Machine Learning Algorithm Cheat Sheet.

Deep learning, machine learning, and AI



Artificial Intelligence

Any technique that enables computers to mimic human intelligence. It includes *machine learning*

Machine Learning

A subset of AI that includes techniques that enable machines to improve at tasks with experience. It includes *deep learning*

Deep Learning

A subset of machine learning based on neural networks that permit a machine to train itself to perform a task.

Consider the following definitions to understand deep learning vs. machine learning vs. AI:

- **Deep learning** is a subset of machine learning that's based on artificial neural networks. The *learning process* is *deep* because the structure of artificial neural networks consists of multiple input, output, and hidden layers. Each layer contains units that transform the input data into information that the next layer can use for a certain predictive task. Thanks to this structure, a machine can learn through its own data processing.

- **Machine learning** is a subset of artificial intelligence that uses techniques (such as deep learning) that enable machines to use experience to improve at tasks. The *learning process* is based on the following steps:

    1. Feed data into an algorithm. (In this step you can provide additional information to the model, for example, by performing feature extraction.)
    2. Use this data to train a model.
    3. Test and deploy the model.
    4. Consume the deployed model to do an automated predictive task. (In other words, call and use the deployed model to receive the predictions returned by the model.)

- **Artificial intelligence (AI)** is a technique that enables computers to mimic human intelligence. It includes machine learning.

By using machine learning and deep learning techniques, you can build computer systems and applications that do tasks that are commonly associated with human intelligence. These tasks include image recognition, speech recognition, and language translation.

Techniques of deep learning vs. machine learning

Now that you have the overview of machine learning vs. deep learning, let's compare the two techniques. In machine learning, the algorithm needs to be told how to make an accurate prediction by consuming more information (for example, by performing feature extraction). In deep learning, the algorithm can learn how to make an accurate prediction through its own data processing, thanks to the artificial neural network structure.

The following table compares the two techniques in more detail:

	All machine learning	Only deep learning
Number of data points	Can use small amounts of data to make predictions.	Needs to use large amounts of training data to make predictions.
Hardware dependencies	Can work on low-end machines. It doesn't need a large amount of computational power.	Depends on high-end machines. It inherently does a large number of matrix multiplication operations. A GPU can efficiently optimize these operations.
Featurization process	Requires features to be accurately identified and created by users.	Learns high-level features from data and creates new features by itself.
Learning approach	Divides the learning process into smaller steps. It then combines the results from each step into one output.	Moves through the learning process by resolving the problem on an end-to-end basis.
Execution time	Takes comparatively little time to train, ranging from a few seconds to a few hours.	Usually takes a long time to train because a deep learning algorithm involves many layers.
Output	The output is usually a numerical value, like a score or a classification.	The output can have multiple formats, like a text, a score or a sound.

What is transfer learning?

Training deep learning models often requires large amounts of training data, high-end compute resources (GPU, TPU), and a longer training time. In scenarios when you don't have any of these available to you, you can shortcut the training process using a technique known as *transfer learning*.

Transfer learning is a technique that applies knowledge gained from solving one problem to a different but related problem.

Due to the structure of neural networks, the first set of layers usually contains lower-level features, whereas the final set of layers contains higher-level features that are closer to the domain in question. By repurposing the final layers for use in a new domain or problem, you can significantly reduce the amount of time, data, and compute resources needed to train the new model. For example, if you already have a model that recognizes cars, you can repurpose that model using transfer learning to also recognize trucks, motorcycles, and other kinds of vehicles.

Learn how to apply transfer learning for image classification using an open-source framework in Azure Machine Learning : Train a deep learning PyTorch model using transfer learning.

Deep learning use cases

Because of the artificial neural network structure, deep learning excels at identifying patterns in unstructured data such as images, sound, video, and text. For this reason, deep learning is rapidly transforming many industries, including healthcare, energy, finance, and transportation. These industries are now rethinking traditional business processes.

Some of the most common applications for deep learning are described in the following paragraphs. In Azure Machine Learning, you can use a model from you build from an open-source framework or build the model using the tools provided.

Named-entity recognition

Named-entity recognition is a deep learning method that takes a piece of text as input and transforms it into a pre-specified class. This new information could be a postal code, a date, a product ID. The information can then be stored in a structured schema to build a list of addresses or serve as a benchmark for an identity validation engine.

Object detection

Deep learning has been applied in many object detection use cases. Object detection comprises two parts: image classification and then image localization. Image *classification* identifies the image's objects, such as cars or people. Image *localization* provides the specific location of these objects.

Object detection is already used in industries such as gaming, retail, tourism, and self-driving cars.

Image caption generation

Like image recognition, in image captioning, for a given image, the system must generate a caption that describes the contents of the image. When you can detect and label objects in photographs, the next step is to turn those labels into descriptive sentences.

Usually, image captioning applications use convolutional neural networks to identify objects in an image and then use a recurrent neural network to turn the labels into consistent sentences.

Machine translation

Machine translation takes words or sentences from one language and automatically translates them into another language. Machine translation has been around for a long time, but deep learning achieves impressive results in two specific areas: automatic translation of text (and translation of speech to text) and automatic translation of images.

With the appropriate data transformation, a neural network can understand text, audio, and visual signals. Machine translation can be used to identify snippets of sound in larger audio files and transcribe the spoken word or image as text.

Text analytics

Text analytics based on deep learning methods involves analyzing large quantities of text data (for example, medical documents or expenses receipts), recognizing patterns, and creating organized and concise information out of it.

Companies use deep learning to perform text analysis to detect insider trading and compliance with government regulations. Another common example is insurance fraud: text analytics has often been used to analyze large amounts of documents to recognize the chances of an insurance claim being fraud.

Artificial neural networks

Artificial neural networks are formed by layers of connected nodes. Deep learning models use neural networks that have a large number of layers.

The following sections explore most popular artificial neural network typologies.

Feedforward neural network

The feedforward neural network is the most simple type of artificial neural network. In a feedforward network, information moves in only one direction from input layer to output layer. Feedforward neural networks transform an input by putting it through a series of hidden layers. Every layer is made up of a set of neurons, and each layer is fully connected to all neurons in the layer before. The last fully connected layer (the output layer) represents the generated predictions.

Recurrent neural network (RNN)

Recurrent neural networks are a widely used artificial neural network. These networks save the output of a layer and feed it back to the input layer to help predict the layer's outcome. Recurrent neural networks have great learning abilities. They're widely used for complex tasks such as time series forecasting, learning handwriting, and recognizing language.

Convolutional neural network (CNN)

A convolutional neural network is a particularly effective artificial neural network, and it presents a unique architecture. Layers are organized in three dimensions: width, height, and depth. The neurons in one layer connect not to all the neurons in the next layer, but only to a small region of the layer's neurons. The final output is reduced to a single vector of probability scores, organized along the depth dimension.

Convolutional neural networks have been used in areas such as video recognition, image recognition, and recommender systems.

Generative adversarial network (GAN)

Generative adversarial networks are generative models trained to create realistic content such as images. It is made up of two networks known as generator and discriminator. Both networks are trained simultaneously. During training, the generator uses random noise to create new synthetic data that closely resembles real data. The discriminator takes the output from the generator as input and uses real data to determine whether the generated content is real or synthetic. Each network is competing with each other. The generator is trying to generate synthetic content that is indistinguishable from real content and the discriminator is trying to correctly classify inputs as real or synthetic. The output is then used to update the weights of both networks to help them better achieve their respective goals.

Generative adversarial networks are used to solve problems like image to image translation and age progression.

Transformers

Transformers are a model architecture that is suited for solving problems containing sequences such as text or time-series data. They consist of encoder and decoder layers . The encoder takes an input and maps it to a numerical representation containing information such as context. The decoder uses information from the encoder to produce an output such as translated text. What makes transformers different from other architectures containing encoders and decoders are the attention sub-layers. Attention is the idea of focusing on specific parts of an input based on the importance of their context in relation to other inputs in a sequence. For example, when summarizing a news article, not all sentences are relevant to describe the main idea. By focusing on key words throughout the article, summarization can be done in a single sentence, the headline.

Transformers have been used to solve natural language processing problems such as translation, text generation, question answering, and text summarization.

Some well-known implementations of transformers are:

- Bidirectional Encoder Representations from Transformers (BERT)
- Generative Pre-trained Transformer 2 (GPT-2)
- Generative Pre-trained Transformer 3 (GPT-3)

Next steps

The following articles show you more options for using open-source deep learning models in Azure Machine Learning:

- Classify handwritten digits by using a TensorFlow model
- Classify handwritten digits by using a TensorFlow estimator and Keras
- Classify handwritten digits by using a Chainer model

Is this page helpful?

Yes No

Recommended content

ML Studio (classic): Deploy a web service - Azure

How to convert a training experiment to a predictive experiment, prepare it for deployment, then deploy it as a Machine Learning Studio (classic) web service.

Machine learning fairness (preview) - Azure Machine Learning

Learn about machine learning fairness and how the Fairlearn Python package can help you assess and mitigate unfairness.

PCA-Based Anomaly Detection: Module reference - Azure Machine Learning

Learn how to use the PCA-Based Anomaly Detection module to create an anomaly detection model based on principal component analysis (PCA).

What is responsible machine learning (preview) - Azure Machine Learning

Learn what responsible machine learning is and how to use it with Azure Machine Learning to understand models, protect data and control the model lifecycle.

Show more