There is a close connection between machine learning and compression. A system that predicts the posterior probabilities of a sequence given its entire history can be used for optimal data compression (by using arithmetic coding on the output distribution). Conversely, an optimal compressor can be used for prediction (by finding the symbol that compresses best, given the previous history). This equivalence has been used as a justification for using data compression as a benchmark for "general intelligence".

An alternative view can show compression algorithms implicitly map strings into implicit feature space vectors, and compression-based similarity measures compute similarity within these feature spaces. For each compressor C(.) we define an associated vector space ℵ, such that C(.) maps an input string x, corresponding to the vector norm ||~x||. An exhaustive examination of the feature spaces underlying all compression algorithms is precluded by space; instead, feature vectors chooses to examine three representative lossless compression methods, LZW, LZ77, and PPM.

According to AIXI theory, a connection more directly explained in Hutter Prize, the best possible compression of x is the smallest possible software that generates x. For example, in that model, a zip file's compressed size includes both the zip file and the unzipping software, since you can not unzip it without both, but there may be an even smaller combined form.

Examples of AI-powered audio/video compression software include NVIDIA Maxine, AIVC.[29] Examples of software that can perform AI-powered image compression include OpenCV, TensorFlow, MATLAB's Image Processing Toolbox (IPT) and High-Fidelity Generative Image Compression.

In unsupervised machine learning, k-means clustering can be utilized to compress data by grouping similar data points into clusters. This technique simplifies handling extensive datasets that lack predefined labels and finds widespread use in fields such as image compression.

Data compression aims to reduce the size of data files, enhancing storage efficiency and speeding up data transmission. K-means clustering, an unsupervised machine learning algorithm, is employed to partition a dataset into a specified number of clusters, k, each represented by the centroid of its points. This process condenses extensive datasets into a more compact set of representative points. Particularly beneficial in image and signal processing, k-means clustering aids in data reduction by replacing groups of data points with their centroids, thereby preserving the core information of the original data while significantly decreasing the required storage space.

Large language models (LLMs) are also capable of lossless data compression, as demonstrated by DeepMind's research with the Chinchilla 70B model. Developed by DeepMind, Chinchilla 70B effectively compressed data, outperforming conventional methods such as Portable Network Graphics (PNG) for images and Free Lossless Audio Codec (FLAC) for audio. It achieved compression of image and audio data to 43.4% and 16.4% of their original sizes, respectively.