# Terminology

1. Distillation：

a technique used in machine learning to transfer knowledge from a large, complex model (teacher model) to a smaller, more lightweight model (student model). The primary goal of knowledge distillation is to distill the knowledge or information learned by the teacher model into a more compact and computationally efficient student model.

1. Dense features

"Dense features" typically refer to high-dimensional feature representations in machine learning and deep learning models. These features are dense in the sense that they are represented by a large number of values, often in the form of continuous or real numbers, as opposed to sparse features that are represented by a smaller set of non-zero values.

1. Contrastive learning

is a type of self-supervised learning technique that aims to learn useful representations by maximizing the similarity between similar pairs of data points while minimizing the similarity between dissimilar pairs. In contrastive learning, the model learns by comparing and contrasting pairs of data points in a way that similar data points are brought closer together in the learned representation space, while dissimilar data points are pushed farther apart.

1. UMAP

UMAP (Uniform Manifold Approximation and Projection) is a dimensionality reduction technique that is widely used for visualizing high-dimensional data in a lower-dimensional space. UMAP is particularly effective for preserving the local and global structure of the data, making it a popular choice for exploratory data analysis, clustering, visualization, and feature engineering tasks.

1. In-context learning (ICL)

Learns a new task from a small set of examples presented within the context (the prompt) at inference time. LLMs trained on sufficient data exhibit ICL, even though they are trained only with the objective of next token prediction. Much of the interest in LLMs is due to the prompting with examples as it enables applications on novel tasks without the need for fine-tuning the LLM.

1. Readout: just the last layer of the network to change internal representations to meaningful format (eg. probability)

A readout in deep learning refers to the final layer or mechanism that converts the learned representations into the desired output format. It's essentially the interface between the internal network representations and the final prediction or output.

1. Noise ceiling

estimate how much stimulus-related variance you have in the data, (and the rest is noise)

assumption underlying this is that stimulus responses are fixed, and differ only in noise

for instance, if you have a correlation between data and model of 0.3 but your noise ceiling estimate is 0.4, you obtain a 'corrected' correlation of 0.3/0.4=0.75

# sNormalization

(<https://medium.com/@sachinsoni600517/layer-normalization-in-transformer-1a2efbff8b85>)

I. def: normalization in deep learning refers to the process of transforming data to conform to specific statistical properties.

II. place to use: Input Data / Hidden Layer Activations (layer norm: stabilize and accelerate training, especially in deep networks)

II. common types

1. standardization: each data adjusted by - mean of its column then / standard deviation

2. min-max normalization: data is scaled to fit within a given range.

3. Batch norm:

Bg: imagine a neural network with multiple hidden layers. During training, the distribution of activations in these layers can change because the weights are updated. This shifting distribution makes it challenging for the network to learn effectively, leading to unstable training. (Internal covariate shift:  distribution of activations changes during training due to the constant updates to the network’s weights)

Detail: normalizing the activations within each layer, ensuring they follow a consistent distribution with a mean of zero and a standard deviation of one.