# AUTOENCODERS

Before deep learning, we had to manually extract features meaning we told the computer *what* to look for:

For images: edges, corners, colors
For text: frequency of words
For signals: wave amplitudes

But there was a problem —
👉 Manual features missed what truly mattered.
👉 Data kept getting bigger: images, audio, video.

So, researchers asked:
💡 “Can a neural network *learn* the key features on its own?”

That’s where **autoencoders** came in.

An autoencoder is a neural network that automatically learns to compress data and reconstruct it back.

.

## ➢ The Structure of an Autoencoder

It’s always made of three parts:
Encoder: Compresses input → small latent vector
Latent space: The compressed data representation
Decoder: Reconstructs the input from that latent vector

**Think of it as: Input → Encoder → Latent (compressed) → Decoder → Output**

## ➢ How It Works — Step by Step

Let’s take an image example to visualize.

**Step 1: Input**
Give the network an image (e.g., 28×28 → 784 pixels).

**Step 2: Encoder**
Compress the input into a small vector (e.g., 32 numbers).

**Step 3: Decoder**
Reconstruct the image from that small vector.

**Step 4: Train**
Compare the reconstruction to the original, compute the loss, update the network, and repeat until the reconstruction is good.

## ➢ Need for Autoencoders

- **Dimensionality Reduction:** Compress large data into smaller, meaningful representations.
- **Noise Removal:** Denoise images or signals.
- **Feature Learning:** Automatically learn important features for tasks like classification.
- **Data Generation:** Create new samples similar to the input (e.g., images).

## ➢ Real World Examples

| Domain | Example |
|---|---|
| **Image** | Remove noise from photos or compress images efficiently |
| **Finance** | Detect fraud by spotting abnormal transactions |
| **Healthcare** | Find unusual patient scans or medical readings |
| **Manufacturing** | Detect machine faults using sensor patterns |
| **Text / NLP** | Represent sentences as compact vectors (semantic meaning) |

## ➢ Types of Autoencoders

- **Vanilla Autoencoder**
  **Use:** Basic dimensionality reduction and feature learning.

- **Denoising Autoencoder**
  **Use:** Remove noise from images or signals.

- **Sparse Autoencoder**
  **Use:** Learn important/rare features by enforcing sparsity.

- **Variational Autoencoder (VAE)**
  **Use:** Generate new data similar to training data

- **Convolutional Autoencoder**
  **Use:** Image compression, reconstruction, and processing.

- **Stacked Autoencoder**
  **Use:** Deep feature learning by stacking multiple layers.

```
from tensorflow.keras import layers, models

# Simple autoencoder
autoencoder = models.Sequential([
    layers.Dense(32, activation='relu', input_shape=(784,)),  # Encoder
    layers.Dense(784, activation='sigmoid')                   # Decoder
])
autoencoder.compile(optimizer='adam', loss='mse')
```