

What is a Neural Network?

A **Neural Network** is a computer system designed to learn patterns from data in a way that is loosely inspired by how the human brain works.

In simple words:

A neural network is a system that learns by **adjusting connections between many small decision units called neurons**.

Basic idea (very simple)

Think of it like this:

- You give input (data)
- The network processes it through many small steps
- It produces an output (prediction)

Example:

Image of a cat → output: “Cat”

Structure of a Neural Network

A neural network is made of **layers**:

Input Layer → Hidden Layers → Output Layer

1. Input Layer

- Takes raw data
- Example:
 - image pixels
 - text numbers
 - sensor values

This is just the starting point.

2. **Hidden Layers (most important part)**

These layers do the actual “thinking”.

Each layer contains many small units called:

Neurons

What a neuron does:

Each neuron:

1. Receives inputs
 2. Multiplies them with weights
 3. Adds bias
 4. Applies activation function
 5. Sends output to next layer
-

Simple analogy:

Imagine decision-making:

- Friend 1 checks shape
- Friend 2 checks color
- Friend 3 checks size

Then they combine opinions to decide:

“Is this a cat?”

That’s what hidden layers do.

3. **Output Layer**

This gives final result:

Examples:

- “Cat” / “Dog”
- Spam / Not spam

- Price prediction

One-line summary

A neural network is a system of interconnected “neurons” that learn patterns from data by adjusting weights through repeated training and error correction.

What is Transformer Architecture?

A **Transformer architecture** is a type of neural network designed to understand and generate sequences of data (like text) very efficiently.

It is the foundation of modern AI systems like **ChatGPT, Google Translate, and many LLMs.**

□ Main idea of Transformer

A Transformer has one powerful concept:

□ **Self-Attention**

It allows the model to decide:

“Which words in this sentence matter most for understanding each word?”

Simple example

Sentence:

“The cat that I saw yesterday was hungry”

When understanding “was”, the model focuses on:

- cat ☑ (important)
- hungry ☑
- ignores less important words

☐ It connects meaning across the whole sentence.

How Transformer works (step-by-step)

1. Input (Text)

Example:

“I love AI”

2. Tokenization

Text is broken into tokens:

["I", "love", "AI"]

3. Embedding (numbers)

Each word becomes numbers (vectors):

- “love” → [0.2, 0.8, ...]
 - “AI” → [0.9, 0.3, ...]
-

4. Self-Attention layer

Each word asks:

“Which other words in this sentence are important to me?”

This creates relationships between words.

5. Multiple layers

Transformer has many layers:

- First layers: basic patterns (grammar)
 - Middle layers: meaning
 - Deep layers: context + reasoning
-

6. Output layer

Finally:

predicts next word or produces output

Key component: Self-Attention (core idea)

For every word, the model calculates:

- Query (what I need)
- Key (what I offer)
- Value (actual meaning)

Then it decides:

how much attention to give each word

Neural networks come in different **types depending** on what kind of data they handle **and what problem they solve** (images, text, sequences, etc.).

Here are the **main types of neural networks explained simply**

1. Feedforward Neural Network (FNN)

What it is:

The simplest type of neural network where data flows in **one direction only**:

Input → Hidden Layers → Output

No loops, no memory.

Example:

- Predicting house price
 - Simple classification (spam/not spam)
-

Key idea:

“Straight-line thinking”

2. Convolutional Neural Network (CNN)

What it is:

A neural network designed for **images and visual data**.

How it works:

It looks for patterns like:

- edges
 - shapes
 - textures
 - objects
-

□ **Example:**

- Face recognition [?](#)
 - Self-driving cars [?](#)
 - Medical scan analysis [?](#)
-

□ **Key idea:**

“Understanding images like human vision”

3. Recurrent Neural Network (RNN)

□ **What it is:**

A neural network designed for **sequences**, where order matters.

It has a form of memory.

□ **Example:**

- Text prediction
 - Speech recognition
 - Time series (stock prices)
-

□ **Problem:**

- Struggles with long sentences
 - Forgets earlier context
-

□ **Key idea:**

“Remembers past information”

4. LSTM (Long Short-Term Memory)

□ **What it is:**

An improved version of RNN that solves memory problems.

□ **Example:**

- Chat systems
 - Language translation
 - Voice assistants
-

□ **Key idea:**

“Better memory for long sequences”

5. GRU (Gated Recurrent Unit)

□ **What it is:**

A simpler and faster version of LSTM.

□ **Example:**

- Mobile AI models [?](#)
- Speech/text tasks

□ **Key idea:**

“Fast memory-based learning”

6. Transformer Neural Network [?](#) (Modern AI)

□ **What it is:**

A powerful architecture that uses **self-attention instead of memory loops**.

□ **Example:**

- ChatGPT [?](#)
 - Google Translate [?](#)
 - Gemini, Claude, etc.
-

□ **Key idea:**

“Understands relationships between all words at once”

Simple comparison

FNN → simple predictions
CNN → images
RNN → sequences (basic)

LSTM/GRU → improved sequence memory
Transformer → modern AI (best for language)

Quick summary table

| Type | Best for | Key feature |
|-------------|----------------|--|
| FNN | Simple data | No memory |
| CNN | Images | Detects patterns |
| RNN | Sequences | Has memory |
| LSTM | Long sequences | Better memory |
| GRU | Fast sequences | Efficient memory |
| Transformer | Text + AI | Attention mechanism <input type="checkbox"/> |

One-line summary

Neural networks have different types like CNN, RNN, LSTM, and Transformers, each designed for specific tasks such as images, sequences, or language understanding.