SkillNavora · Written by SkillNavora · Plain-English guide · Fake data only

NLP, Neural Networks and LPUs, Explained Simply

Three words come up again and again in AI: **natural language processing (NLP)**, **neural networks** and **LPU**. This guide explains each one in plain English, shows how they fit together, and links to the deeper lessons on this site. You do not need any background.

> **The short version.** A **neural network** is a program that learns patterns from examples. **NLP** is the job of making computers work with human language. An **LPU** is a kind of chip built to run language models quickly. Most modern language AI uses all three ideas together.

Part 1. Neural networks in one picture

A neural network is built from many small units called **neurons**. Each neuron takes numbers in, multiplies each one by a **weight**, adds them up, and passes the result on. Learning means adjusting the weights until the network's answers match the examples.

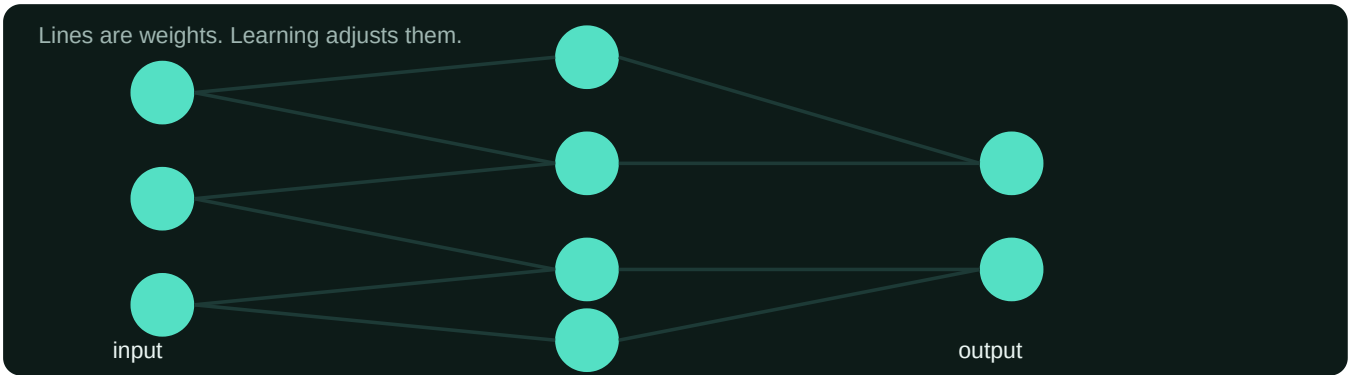


Figure 1. Input, hidden layer and output. Each line carries a number that learning adjusts.

Three words to know

Word	Plain meaning
Weight	How strongly one neuron's signal affects the next
Layer	A group of neurons that work at the same stage
Training	Adjusting weights many times until answers improve

Deep learning just means a neural network with many layers. Lesson ML-26 on this site builds one step by step.

Part 2. Natural language processing: teaching computers to read

Human language is messy. The same idea can be written many ways, and meaning depends on context. NLP is the set of methods that turn text into something a computer can use.

NLP task	What it does	Example
Cleaning	Removes noise such as extra spaces and symbols	"Great!!! product" becomes "great product"
Tokenising	Splits text into pieces (tokens)	"I love it" becomes 3 tokens
Sentiment	Decides if text is positive or negative	"The app crashes" becomes negative
Classification	Puts text into groups	A complaint goes to "billing"
Summarising	Shortens text and keeps the key points	A 10-page policy becomes 5 bullets
Question answering	Finds or writes an answer from text	"How many leave days?" answered from a policy

From words to numbers

A computer cannot use words directly. It turns each token into a list of numbers called an **embedding**. Words with similar meaning get similar numbers. Earlier methods counted words (for example TF-IDF, which gives more weight to words that are rare across documents but common in one). Modern language models learn their embeddings from very large amounts of text.

Lesson ML-25 on this site walks through cleaning, TF-IDF and sentiment classification with working code.

Part 3. How NLP and neural networks meet

Most modern NLP uses neural networks. A language model is a large neural network that predicts the next token. It learns this by reading very large amounts of text and adjusting its weights each time it guesses wrong.

1. **Input:** text is split into tokens and turned into embeddings.
2. **Processing:** many layers combine the embeddings, so each token takes context from the others.
3. **Output:** a score for every possible next token. The model picks one (see the tokens guide on sampling).
4. **Repeat:** the chosen token is added to the input, and the process runs again for the next token.

Part 4. What an LPU is

An **LPU** (language processing unit) is a type of processor designed to run language models quickly. Groq is one company that uses the name for its chips. Most AI today runs on **GPUs** (graphics processors), which are good at many kinds of parallel maths. An LPU is designed around the sequential way language models generate tokens one after another.

Word	What it is	Why you care
CPU	General-purpose processor in every computer	Runs everyday software; slow for large models
GPU	Processor with thousands of small cores	The most common hardware for training and running AI
LPU	Chip designed for fast language model output (for example Groq's)	Some providers offer very fast responses on it

> **Honesty note.** Chip speeds, availability and prices change quickly and are often quoted by the company that sells them. Check the provider's current documentation before you rely on any speed or cost claim. This guide does not state any LPU performance figure.

Why it matters for a product manager: faster responses can change what users will accept. Speed, cost and quality trade off against each other, so the choice of hardware and model is a product decision, not only an engineering one.

Part 5. A quick map of the terms

Term	Layer	One-line meaning
NLP	Goal	Making computers work with human language
Neural network	Method	Learning patterns with weighted layers of neurons
Language model	Product	A neural network trained to predict and write text
LPU / GPU / CPU	Hardware	The chip that runs the model

Part 6. Worked example: a support-ticket sorter

1. **Goal:** sort fake support tickets into Billing, Bug or Other.
2. **Classic NLP:** clean the text, turn it into TF-IDF numbers, train a simple classifier on 50 labelled fake tickets. Measure accuracy on 20 held-out tickets.

3. **Language model:** ask the model to choose one label from the list and give a one-line reason. Lock the output to the three labels.
4. **Compare:** speed, cost per ticket (see the tokens guide), and accuracy on the same 20 tickets.
5. **Choose:** the simplest method that meets your accuracy target. Use a language model where the rules are hard to write.

Pass when: you can explain which method you chose, what it costs per ticket, and what accuracy it reached on the same test set.

Part 7. Common mistakes

Mistake	Better habit
Using a large model where a simple rule works	Start with the simplest method; measure before you upgrade
Judging speed from a marketing claim	Time the same 20 requests yourself
Confusing the model with the chip	Name both separately in every plan
Testing only on clean sentences	Include typos, slang and mixed languages

Part 8. Check yourself

1. What does a weight do in a neural network?
2. Name one NLP task that is not a language model.
3. What is an embedding?
4. What does LPU stand for, and what is it designed for?
5. Why should you test speed yourself?

Answers: 1. It sets how strongly one signal affects the next; learning adjusts it. 2. Sentiment analysis, classification or cleaning. 3. A list of numbers that represents the meaning of a token. 4. Language processing unit; it is designed to run language models quickly. 5. Published claims change and may not match your use.

Part 9. Glossary

Term	Plain meaning
NLP	Natural language processing: computers working with human language.
Neural network	A program that learns patterns through weighted layers.

Weight	A number learning adjusts to change a network's output.
Embedding	Numbers that stand for the meaning of a token.
TF-IDF	A word-weighting method for documents, used in classic NLP.
LPU	Language processing unit: a chip designed for fast language model output.
GPU	Graphics processor, widely used to run and train AI models.

Related lessons on SkillNavora: ML-25 (NLP and text analytics), ML-26 (neural networks and deep learning), GenAI-01 (how large language models work).