CAMBIUM ACADEMY · COURSE 01

Intro to AI

How AI works, how it is trained, and which model to use



Free forever. No signup. **cambium ai research institution**

July 2026 edition. The AI world moves fast, so facts carry dates.

WELCOME

This course was built for beginners

No background needed
If you can use a web browser, you are qualified.

No math, no code
We build intuition with pictures and plain English.

Jargon gets translated
Every technical term arrives with a one-line meaning.

WHAT YOU NEED

- A web browser
- About 90 minutes
- Zero dollars

OUTCOMES

Five things you will walk away with

Explain AI in one sentence
And know how AI, machine learning, and generative AI relate.

See how models really work
Tokens, neural networks, and the transformer, minus the math.

Understand training
From a mountain of text to a helpful assistant, in four steps.

Know the 2026 landscape
Who makes the top models and what makes each one special.

Pick the right model
A practical cheat sheet for writing, coding, research, and more.

COURSE MAP

Six short modules, one quiz, one certificate

1 **What is AI?**
10 min

2 **How models work**
12 min

3 **How models are trained**
12 min

4 **The 2026 model landscape**
12 min

5 **Which model for which job**
8 min

6 **Using AI well**
8 min

Then: flashcards and the hands-on AI Lab unlock the 20-question quiz. Score 70% or better and print your certificate on the spot.

MODULE 1

What is AI?

Software that learns from examples instead of following hand-written rules.



MODULE 1 · DEFINITION

AI in one sentence

*Artificial intelligence is software that **learns patterns from examples** and uses them to make useful guesses about things it has never seen.*

Traditional software

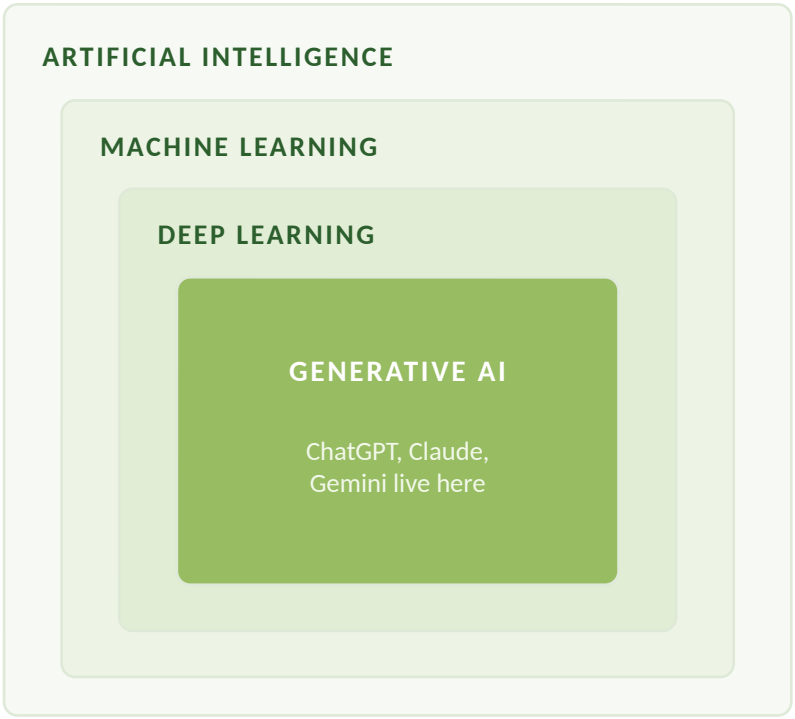
A person writes every rule by hand. "If the email says WIN CASH, mark it as spam." Breaks the moment reality gets messy.

AI / machine learning

Show it a million spam emails and a million real ones. It finds the telltale patterns itself, including ones no human ever wrote down.

MODULE 1 · THE FAMILY TREE

AI is the family. Generative AI is the newest child.



Artificial intelligence
Any technique that makes machines act smart, from chess engines to chatbots.

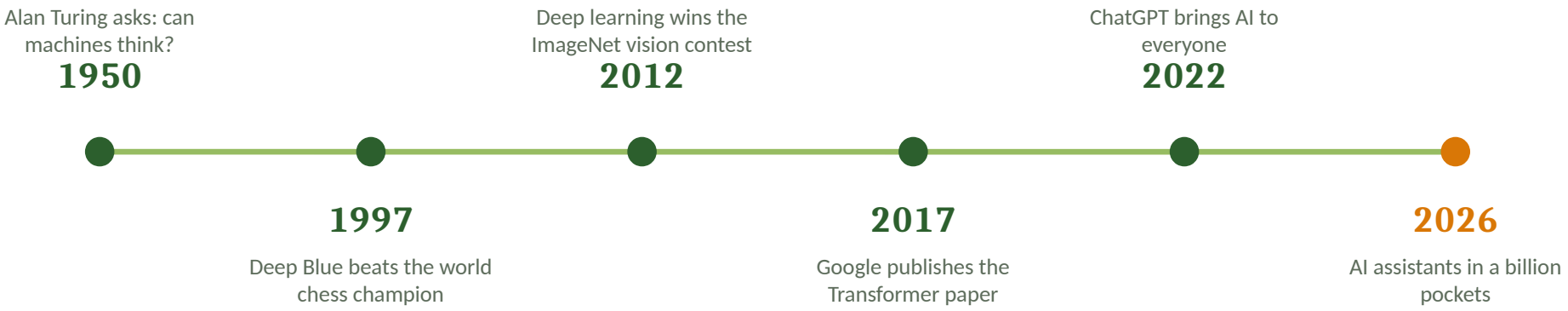
Machine learning
The dominant technique: learn patterns from data instead of hand-written rules.

Deep learning
Machine learning with many-layered neural networks. Took off in 2012.

Generative AI
Deep learning that creates new text, images, audio, and video. The 2022 breakthrough.

MODULE 1 · A SHORT HISTORY

Seventy-five years in one slide



The lesson: AI is not new. What changed recently is scale, data, and a 2017 invention called the transformer.

MODULE 1 · ALREADY EVERYWHERE

You used AI before breakfast

Spam filters

ML has guarded your inbox for 20 years

Phone cameras

Every photo is cleaned up by a neural net

Maps and traffic

Route predictions learned from millions of trips

Autocomplete

Your keyboard predicts the next word, LLM-style

Recommendations

Netflix, YouTube, and Spotify picks

Voice assistants

Speech to text is deep learning too

Chatbots are simply the newest face of a technology you already trust daily.

MODULE 2

How AI models work

A language model is a prediction machine.



MODULE 2 · THE BIG IDEA

One trick underneath it all: predict the next word

"The capital of France is ___"

Paris 97%
Lyon 1% · a 0.5% · …

Repeat the trick
Predict a word, add it to the sentence, predict again. A paragraph is thousands of predictions in a row.

Trained on trillions of words
Good prediction quietly forces the model to absorb grammar, facts, style, and reasoning patterns.

That is the whole magic
Essays, code, and advice all emerge from one skill: guessing the next word extremely well.

MODULE 2 · TOKENS

AI does not read words. It reads tokens.

Under	stand	ing	to	kens
16118	2752	278	311	12456

one word can be several tokens, and every token is just a number to the model

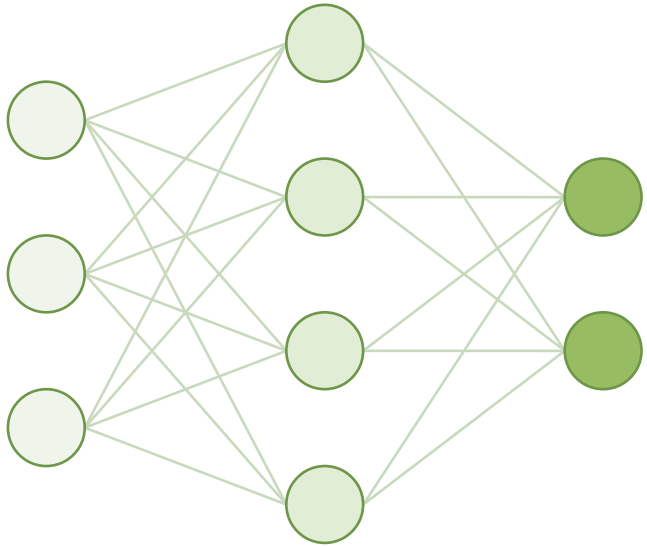
A token is a chunk of text
Roughly three quarters of an English word. 1,000 tokens is about 750 words.

Models have a token budget
The context window. Fill it up and the oldest content falls out of memory.

Tokens are the currency of AI
API pricing, speed ratings, and limits are all counted in tokens.

MODULE 2 · NEURAL NETWORKS

Billions of tiny dials, tuned by practice



Every connection is a dial

A weight: one number that gets nudged up or down during training.

Learning = tuning the dials

Wrong guess? Nudge billions of weights so the right answer scores higher next time.

Scale is the surprise

Frontier models have hundreds of billions of weights. Somehow, fluency emerges.

MODULE 2 · THE TRANSFORMER

The 2017 invention that changed everything

"The trophy did not fit in the suitcase because it was too big." Which "it"? ***Attention links "it" to "trophy".***

Attention: every word looks at every other word

The model learns which words matter to which, no matter how far apart they sit.

It reads everything at once

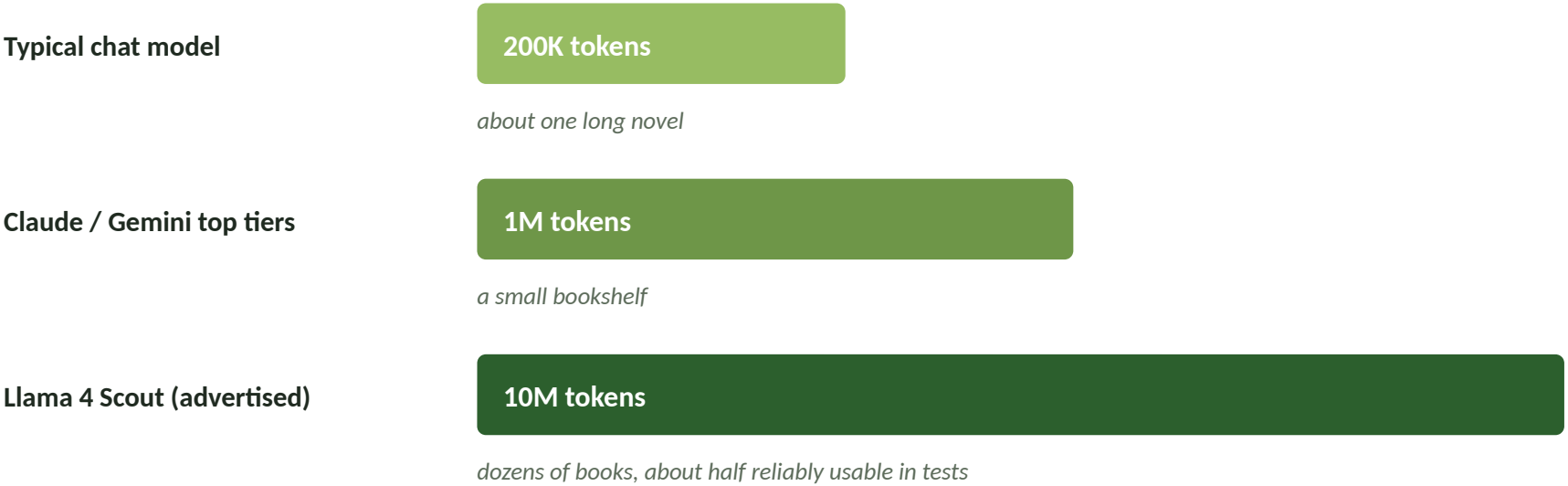
Older AI read word by word. Transformers process a whole text in parallel, so training on internet-scale data became affordable.

It is the T in GPT

Generative Pre-trained Transformer. Claude, Gemini, and Llama are transformers too.

MODULE 2 · WORKING MEMORY

The context window: how much it can hold in mind



When the window fills, the earliest content silently falls out. If a long chat gets forgetful, start a fresh one and restate what matters.

MODULE 3

How models are trained

Read everything. Learn the job. Learn manners.



MODULE 3 · THE RECIPE

From raw internet to helpful assistant, in four steps

1

Gather data

Trillions of words: web pages, books, code, articles, transcripts.

>

2

Pretrain

Predict the next token, billions of times. Absorbs language and facts.

>

3

Fine-tune

Teach it to follow instructions and hold conversations.

>

4

Human feedback

People rank answers. The model learns what helpful and safe look like.

Steps 1 and 2 build a brilliant autocomplete. Steps 3 and 4 turn it into an assistant.

MODULE 3 · THE PRICE TAG

Why only a handful of labs train frontier models

Months

of continuous training for one frontier run

10,000s

of specialized GPU chips running in parallel

$100M+

estimated cost of a single frontier training run

The efficiency race is on. In 2026, open models like DeepSeek V4 match top-tier performance at roughly one tenth the serving cost per token, proof that clever engineering can beat raw spending.

MODULE 3 · WHY CHATBOTS FEEL HUMAN

Same brain, before and after finishing school

BASE MODEL

A brilliant autocomplete with no manners.

```
You: How do I bake bread?
Base model: asked Jim, who had baked for
years...
```

It just continues the text. It was never taught to answer.

AFTER FINE-TUNING + FEEDBACK

A helpful assistant.

- ✓ Answers the actual question, step by step
- ✓ Declines harmful requests
- ✓ Admits uncertainty instead of bluffing

This stage is why talking to AI feels natural.

MODULE 3 · THE CATCH

Why AI confidently makes things up

It always answers with the most plausible words
Plausible is not the same as true. There is no built-in fact checker.

Gaps get filled smoothly
Ask about something it does not know and it may invent names, dates, citations, or numbers that merely sound right.

Modern tools fight back
Web search, citations, and reasoning modes reduce hallucinations. They do not eliminate them.

The one rule to remember: trust, but verify. Check every name, number, date, and citation before you rely on it.

MODULE 4

The 2026 model landscape

A handful of labs, two philosophies.



MODULE 4 · THE BIG SPLIT

Closed weights or open weights

Closed weights

You use the model as a service. The recipe stays secret.

GPT, Claude, Gemini, Grok
Usually the strongest, easiest to use, and safest defaults. You need internet and their servers.

Open weights

You can download the model and run it yourself.

Llama, DeepSeek, Qwen, Mistral
Free, private, customizable, runs offline. You bring the hardware and the elbow grease.

MODULE 4 · THE FRONTIER THREE

OpenAI, Anthropic, Google

OpenAI · GPT

GPT-5.5 now, GPT-5.6 rolling out July 2026

The most-used AI on earth. New chips serve replies at up to 750 tokens a second, and an Ultra mode puts teams of AI agents on one hard task.

Free: ChatGPT free tier

Anthropic · Claude

Sonnet 5 free, Opus 4.8, new Fable 5 on top

Known for careful long-form work, top-tier coding via Claude Code, and million-token context. Sonnet 5 became the free default on July 1.

Free: claude.ai

Google · Gemini

Gemini 3.1 Pro and 3.5 Flash

Woven through Search, Docs, and Gmail, with the broadest free tier: images, voice, and about five Deep Research reports a month at no cost.

Free: Gemini app

MODULE 4 · MORE HEAVYWEIGHTS

Meta, xAI, DeepSeek, Mistral

Meta

Muse Spark is its new closed flagship, a strategy shift. Open Llama 4 remains, with a 10M-token context on Scout. Free inside WhatsApp and Instagram.

xAI · Grok

Grok 4.3 taps the live X firehose for real-time answers, takes video input, and holds a million tokens. Free tier is thin; most features are paid.

DeepSeek

Open V4 model matches top-tier benchmarks at roughly a tenth of the serving cost. Generous free app. Hosted in China, so mind where your data goes.

Mistral

Europe's champion. Mistral Large 3 is the largest fully open (Apache 2.0) model from a Western lab. Free casual use via its Vibe chat app.

MODULE 4 · THE OPEN-WEIGHTS WAVE

The strongest free models now come from open labs

Qwen 3.5 / 3.6

Alibaba

Open family in every size from 0.8B to 397B parameters. Excellent multilingual skills. Free chat and free weights.

Kimi K2.6

Moonshot AI

Around a trillion parameters, open, and a favorite for agentic coding on a budget. Free chat at kimi.com.

GLM-5.2

Z.ai

MIT licensed with a million-token context, rivaling Western flagships on coding benchmarks. Free chat at chat.z.ai.

Worth knowing: these labs are China-based. The models are world class and free, but chats on their hosted apps fall under Chinese data rules. For sensitive work, run the open weights locally instead.

MODULE 4 · THE SCALE OF IT

Four numbers from 2026

1B+

monthly ChatGPT app users by June 2026, the fastest app ever to get there

900M

monthly users of Google's Gemini app by May 2026

5B

images created with Gemini's Nano Banana tool in about two months

$965B

Anthropic's May 2026 valuation, briefly the highest of any AI lab

Sources and dates for every figure are on the course resources page.

MODULE 4 · OPEN WEIGHTS, YOUR MACHINE

You can run a real AI model on your own computer

Install Ollama, pick a model
Ollama is a free app for Windows, Mac, and Linux that downloads and runs open models with one command.

Everything stays local
No internet needed after download. Nothing you type leaves your machine. Cost: zero, forever.

Small models got good
Qwen3 4B or 8B suits most laptops. Gemma 3 and Phi-4-mini shine on lighter hardware.

Honest trade-off: local models are noticeably weaker than the frontier. Perfect for private notes, drafts, and learning. Not for your hardest problems.

MODULE 5

Which model for which job?

There is no single best model. There is a best model for each job.



MODULE 5 · CHEAT SHEET, PART 1

Words and work

Your job	Reach for	Also great
Everyday writing and Q&A	ChatGPT, free tier	Claude Sonnet 5, often the more natural writer
Coding	Claude with Claude Code	DeepSeek V4 or Kimi K2.6, strong and free
Research with sources	Perplexity, free, cites as it answers	Gemini Deep Research, about 5 free reports a month
Long documents	Claude, million-token context	Gemini 3.1 Pro, 1M context plus Docs and Drive

Snapshot: July 2026. Free tiers and model names shift often; the course page keeps the current version.

MODULE 5 · CHEAT SHEET, PART 2

Media, privacy, and the best free picks

Your job	Reach for	Also great
Images	Gemini's Nano Banana, free	Midjourney, paid, for fine art control
Video clips	Google Veo 3.1, 4K with sound	Kling 3.0, cheaper per clip
Voice and speech	ElevenLabs, free tier	ChatGPT voice mode, simple and bundled
Private and offline	Qwen3 via Ollama, local	Gemma 3 or Phi-4-mini on light hardware
Best free overall	Gemini free tier, broadest features	ChatGPT free tier, easiest to get help with

MODULE 5 · A 10-SECOND DECISION

Four questions beat any leaderboard

Is the data sensitive?
Run an open model locally, or use a paid plan with training turned off.

Do you need checkable sources?
Use a research tool that cites: Perplexity or Deep Research.

Is the input huge?
Pick a million-token context model: Claude or Gemini.

Just everyday help?
Any of the big three free tiers is excellent. Stop optimizing.

Best advice: pick one model and learn it deeply. Prompting skill transfers everywhere; tool-hopping does not compound.

MODULE 6

Using AI well

The human stays in charge.



MODULE 6 · HABITS

Five golden rules

Verify before you rely
Check names, numbers, dates, and citations. AI is confident even when wrong.

Guard private data
Do not paste passwords, client records, or secrets into tools you have not vetted.

You own the judgment
AI drafts, suggests, and accelerates. Decisions and responsibility stay with you.

Disclose when it matters
School, work, publishing: say when AI helped. Honesty ages well.

Reassess monthly
Models leapfrog each other constantly. Today's winner is next month's runner-up.

MODULE 6 · A TASTE OF COURSE 02

Better answers in four moves

Role

Who should the AI be?

Context

What does it need to know?

Task

Exactly what do you want?

Format

How should the answer look?

EXAMPLE PROMPT

You are a seasoned travel planner. Two adults are visiting Boise for three days in October on a modest budget. **Plan a day-by-day itinerary with one outdoor highlight per day.** Present it as a simple table.

Role and task in green, context and format in black. Course 02 goes deep on this next week.

MODULE 6 · KEEP GOING, FREE

Hand-verified free resources

How Machines Learn, CGP Grey
The best 5-minute mental model on YouTube

Intro to Large Language Models, Andrej Karpathy
A legendary 1-hour talk by a top AI researcher

Neural Networks, 3Blue1Brown
Beautiful visual series on what is inside the box

Elements of AI, University of Helsinki
The classic free intro course, 2M+ learners

AI for Everyone, Andrew Ng
Free-to-audit course made for non-engineers

Every link, with time estimates and levels, lives on the course resources page. Each one was verified live in July 2026.

YOUR TURN

Finish the path. Earn your certificate.

- Step 1, the lecture: done, you just finished it
- Step 2: clear all 24 flashcards until they stick
- Step 3: explore the three stations in the AI Lab
- The quiz unlocks: 20 questions, 70% to pass, certificate prints on the spot

Next week · Course 02: Prompting Essentials

Cambium Academy · free forever · new course every week