

How AI Models Work: From Data to Answers, Without the Maths

What AI and machine learning really are, how a model learns from data, how it makes a prediction, the kinds of models you meet every day, and why they sometimes get things badly wrong.

20 min read · Content last verified 8 October 2026 · Reviewed by Shravya M

Free awareness guide. Not legal or financial advice. <https://www.shravyadigital.com/guides/how-ai-models-work>

CONTENTS

1. Before you start
2. Chapter 1: What "AI" actually means
3. Chapter 2: Learning from data
4. Chapter 3: What a model is
5. Chapter 4: Using a model: inference
6. Chapter 5: Different kinds of AI models
7. Chapter 6: Why AI gets things wrong
8. Chapter 7: AI around you
9. Key words

Before you start

This guide is the reading for **Course 1** of the AI Foundations Pathway. It takes about 20 minutes to read. You need no maths and no coding: just curiosity.

By the end you'll be able to explain, in your own words, what an AI model is, how it learns, how it makes a prediction, and why it sometimes gets things badly wrong. Everything else in the pathway (LLMs, RAG, agents) builds on these ideas.

All names and examples here are made up.

Chapter 1: What "AI" actually means

Artificial intelligence (AI) is a broad name for computer systems that do things we'd normally say need human judgement: recognising a face, understanding speech, translating a sentence, or spotting a suspicious payment.

Most AI you meet today is **machine learning**. Instead of a programmer writing every rule by hand ("if the amount is over ₹50,000 and the city is new, flag it"), the computer is shown many examples and **works out the patterns itself**.

What AI is *not*:

- **Not a mind.** It doesn't understand, feel or want anything. It finds patterns and applies them.
- **Not magic.** Everything it does comes from the data it learned from and the way it was built.
- **Not always right.** It gives its best guess, and guesses can be wrong.

The rule: AI is pattern-matching at enormous scale. Powerful, useful, and never infallible.

Chapter 2: Learning from data

A model learns the way a child learns to tell cats from dogs: by seeing lots of examples.

Suppose Meera's team wants AI that spots spam SMS messages. They collect thousands of messages and **label** each one "spam" or "not spam". This collection is the **training data**.

During **training**, the computer:

1. Looks at a message and makes a guess.
2. Checks the guess against the label.
3. Adjusts itself a tiny bit to be less wrong next time.
4. Repeats this millions of times.

Slowly it learns which words, links and patterns tend to mean spam.

Two things decide how good it becomes:

- **How much data:** more varied examples usually mean better learning.
- **How good the data is:** wrong labels, missing kinds of messages, or one-sided examples teach the model the wrong lessons. People say "*garbage in, garbage out*".

Some models learn from labelled examples like these (**supervised learning**). Others find groups and patterns in data that has no labels (**unsupervised learning**). Some learn by trial and reward, like a game player getting a higher score (**reinforcement learning**).

Chapter 3: What a model is

After training, what's left is the **model**: a very large set of numbers, called **parameters**, that capture the patterns it learned.

Think of the parameters as millions of tiny dials. Training is the process of turning each dial until the model's guesses match the examples as closely as possible. Large modern models have billions of these dials.

The model doesn't store a copy of every training example. It stores the *patterns*. That's why it can handle a message it has never seen before, and also why it can't simply "look up" a fact the way a database does.

Remember: *Training is slow and expensive and happens once (or occasionally). Using the model afterwards is quick and happens every time you ask it something.*

Chapter 4: Using a model: inference

Using a trained model is called **inference**. You give it an **input** and it gives an **output**:

- Input: an SMS. Output: "spam, 94% likely".
- Input: a photo. Output: "this is a cat".
- Input: a question. Output: a written answer.

Notice the "94% likely". Most models produce a **probability**, a measure of how confident they are, not a certainty. The app then decides what to do with it: block the message, show a warning, or let it through.

A model only knows what was in its training data up to the day training stopped. That date is called its **knowledge cut-off**. Ask about something that happened afterwards and it may not know, or worse, may guess.

Chapter 5: Different kinds of AI models

Models are built for different jobs:

Kind	What it does	Everyday example
Classification	Puts something in a category	Spam filter, fraud alert on a card payment
Prediction (regression)	Estimates a number	Delivery time on a food app, traffic on maps
Recommendation	Suggests what you might like	Videos, songs, products "for you"
Recognition	Understands images or speech	Face unlock, voice typing in Hindi
Generative AI	Creates new content	Chatbots writing text, apps making images or voices

Generative AI is the newest and most talked about. Instead of choosing from a fixed list of answers, it *creates* new text, images, audio or video based on patterns it learned. ChatGPT, Gemini and Claude are generative models for text; Course 2 explains them in detail.

Chapter 6: Why AI gets things wrong

Because a model learns patterns from data, its mistakes usually come from that data or from being used where it wasn't trained:

- **Bias.** If the training data is one-sided, so is the model. A loan model trained mostly on city borrowers may judge village borrowers unfairly. A face-recognition model trained mainly on one skin tone may work worse on others.
- **Out-of-date knowledge.** It doesn't know what happened after its cut-off.
- **Overfitting.** A model that memorised its training examples too closely does well on them but badly on new, real-world cases, like a student who memorised answers without understanding.
- **Hallucination.** Generative models can produce fluent, confident text that is simply false, such as made-up facts, quotes or references.
- **Being tricked.** People can craft inputs on purpose to fool a model, such as a scam message worded to slip past a spam filter.

None of these are rare accidents. They're built into how learning from data works, so good AI products add human checks, testing, and ways to report mistakes.

The rule: The more an AI decision matters, whether about money, health, a job or the law, the more a human should check it.

Chapter 7: AI around you

You already use dozens of AI models every day, often without noticing:

- Your **bank and UPI app** score payments for signs of fraud.
- **Maps** predict traffic and arrival times.
- **Your keyboard** predicts your next word, and voice typing turns speech into text.
- **Video and shopping apps** recommend what to watch or buy next.
- **Spam filters** in your email and SMS block junk.
- **Chatbots** answer questions and write drafts.

Knowing how these models work helps you use them well: trust them for what they're good at, double-check them where mistakes are costly, and notice when an "AI" claim is just marketing.

Key words

- **AI:** computer systems doing tasks that normally need human judgement.
- **Machine learning:** AI that learns patterns from examples instead of hand-written rules.
- **Training data:** the examples a model learns from.
- **Model / parameters:** what training produces, a huge set of numbers capturing patterns.
- **Inference:** using a trained model to get an answer.
- **Probability / confidence:** how likely the model thinks its answer is.
- **Knowledge cut-off:** the date a model's training data ends.
- **Bias:** unfair, one-sided results caused by one-sided data.
- **Overfitting:** memorising training examples instead of learning general patterns.
- **Generative AI:** AI that creates new text, images, audio or video.
- **Hallucination:** confident output that is false.

Take the course for what you just read

Take the free 10-question assessment and pass with 8 or more. Every course you pass counts towards a Pathway Certificate issued by Shravya Digital, earned when you pass all five courses in the pathway: <https://digital.shravyaworld.org/courses/how-ai-models-work>

Sources

Checked against these primary sources on 8 October 2026. Rules change: if something here looks out of date, trust the source.

- Google: Machine Learning Crash Course (<https://developers.google.com/machine-learning/crash-course>)
- IBM: What is machine learning? (<https://www.ibm.com/think/topics/machine-learning>)
- NITI Aayog: Responsible AI for All (principles) (<https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf>)