EdSutra

Teaching guide

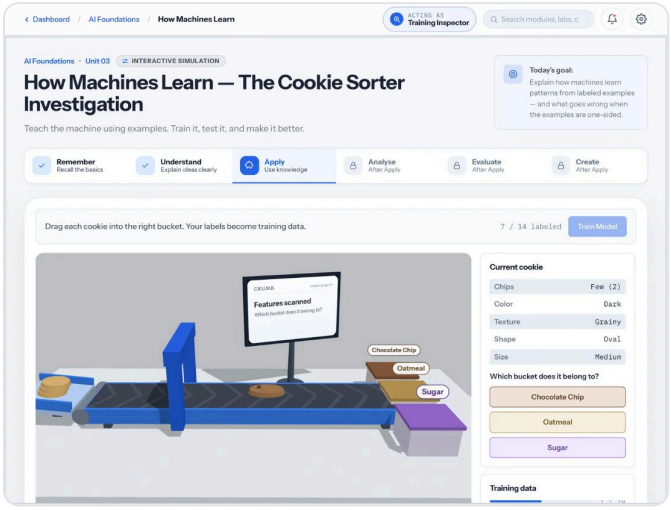
Module 01 · AI Foundations · Unit 03 · Stage 1, Foundations

How Machines Learn

The Cookie Sorter Investigation

Grades 6–9 | **55 minutes** | Role: **Training Inspector** | **Free**, no account needed

Grade 6: About age 11–12 · Year 7 in the UK · Class 6 in India



The big idea

A model learns whatever pattern is easiest in its training examples, which is not always the one we meant. The Cookie Sorter looks perfect, yet on cookies it has never seen it gets about one in four wrong, because it learned color instead of chips. Testing on new examples, and adding counterexamples, is how people catch this and fix it.

The learner's job

Learners are the **Training Inspector** at Halden Institute · Test Bench. The sorter got full marks on its exam. The bakery still gets complaints every week. Their task: find out what the machine really learned, not what it scored.

By the end, learners can say

- I can explain how supervised learning uses labeled examples to learn a pattern
- I can analyze which features a model is using to make predictions
- I can evaluate model behavior with unseen data, accuracy, precision, and recall
- I can diagnose shortcut learning and propose counterexamples that improve generalization
- I can design a stronger feature strategy for a classifier

Before class

- One computer or tablet for each learner or pair, with EdSutra open. Module 1 is free and needs no account.
- Open the studio once on each device before class, so the Unit still runs if the internet drops.
- Play the first two steps yourself, so you know where learners will click.
- Draw two columns on the board, "Training" and "Test". You will fill them in during Analyze.
- Pairs work well: one learner drives, the other is the Training Inspector who notes what the machine leans on.
- Best after AI Alphabet and What is AI?.

Words to introduce

Supervised learning: Learning a pattern from examples that already carry the right answer.

Label: The right answer attached to an example, such as "Chocolate Chip".

Feature: Something the model can measure, such as color or the number of chips.

Training and test data: The examples a model learns from, and the new ones it is checked on.

Shortcut learning: Leaning on an easy pattern that works in training but is not the real reason.

Counterexample: An example added on purpose to break a shortcut.

Accuracy, precision, recall: How often it is right; how many of its picks were right; how many right ones it found.

Running the lesson 55 minutes

Open 5 min

"A machine scored full marks on its exam, but the bakery still gets complaints every week. How can both be true?"

Before anyone clicks, ask: which clue would a machine use to tell a chocolate chip cookie from a sugar cookie? Write the guesses on the board and come back to them in Analyze.

Step	What learners do	Ask the class	Listen for
1. Remember 5 min	Click each part of the sorting machine to learn its name and job, and find all five: input, features, model, decision and buckets.	• Which part turns a cookie into numbers the model can use? • Which part makes the call?	Features are the measurements the model sees, not the cookie itself.
2. Understand 7 min	Press Play and watch cookies get scanned and sorted. Learners notice which features decide each bucket.	• Which feature seemed to decide most of the sorting? • Does the machine know what a cookie is, or only the numbers it was given?	The machine applies a rule to the features it has. It does not understand cookies.
3. Apply 10 min	Learners drag each cookie into the right bucket. Their labels become the training data the model learns from.	• Which cookies were hard to label, and why? • What would happen if one of us labeled carelessly?	Labels are human choices. Mistakes and patterns in them become what the model learns.
4. Analyze 12 min	Run the test batch on cookies the model has never seen. It gets about one in four wrong. Learners inspect each mistake, and can bake their own cookie to try to fool the model.	• What do the wrong cookies have in common? • What did the model lean on: chips, or color? • Look back at our guesses on the board. Were we right?	The model learned color, a shortcut that happened to work in training, instead of chips, the real reason.
5. Evaluate 8 min	Learners read the scorecard (accuracy, precision and recall) and make the call: ship the sorter or send it back.	• Would you ship it? What would the bakery's customers say? • Which number matters most to this bakery, and why?	A decision backed by the test results, not by the training score.
6. Create 8 min	Learners change the features or add new training cookies, retrain, and try to beat Model A.	• Which counterexample did you add, and why did it help? • Did switching a feature off make the model better or worse?	Better features and counterexamples help a model do well on cookies it has never seen.

Steps follow Bloom's revised taxonomy, in the order the Unit presents them.

Deepen the thinking

Critical thinking	A model can be right for the wrong reason. How would you check that a tool marking essays is right for the right reason?
AI thinking	If you could give the sorter one more feature, which would you choose, and what new shortcut might it find?
Ethics and fairness	If a model that approves loans learned a shortcut, who could be harmed, and whose job is it to notice?
Connect to life	Where have you seen a test score that did not show what someone really knew?

Misconceptions to listen for

"If it is accurate on its training data, it is good."

Analyze shows the sorter failing on new cookies. Only test data the model has never seen shows what it really learned.

"The model understands cookies."

It only uses the features it is given. Switch Color off in Create and its decisions change.

"Bias means the machine is being unkind."

Here, bias means learning an easy, wrong pattern from the examples. Counterexamples fix it.

In real life the card learners see at the end of the Unit

- A spam filter learned from emails people sorted into "spam" and "not spam", just like your cookie labels.
- A photo app that finds "dogs" was trained on labeled pictures, then tested on new ones it had never seen.
- At a shop checkout, a card payment can be blocked as "possible fraud" by a model trained on labeled payments.

Try it this week. Sort 10 objects at home into two groups using a secret rule. Ask someone to guess the rule. Did they pick up a shortcut, like color?

Watch out for. A model can be right for the wrong reason, like sorting cookies by color instead of chips. Test it on new examples before you trust it.

Talk about it. Have you ever learned a shortcut that worked in practice but failed on the real test?

Exit ticket

1. What is the difference between training data and test data?

2. Which shortcut did the Cookie Sorter learn, and which clue should it have used?

3. Describe one counterexample that would help, and say why.

Going further

In the full studio, the Neural Playground Lab lets learners train a real network to read handwritten digits and test it on their own writing. For homework, use the "Try it this week" activity from the In real life card.