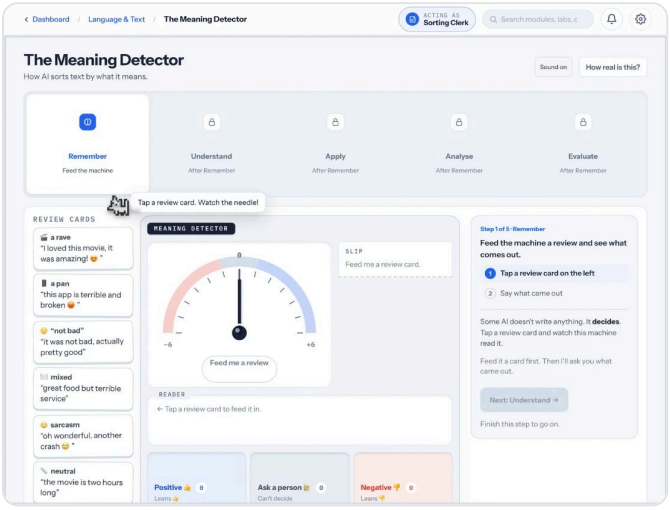
Module 03 · Language & Text · Unit 03 · Stage 2, Capabilities

The Meaning Detector

How AI sorts text by what it means

Grades 7–10 | 20 minutes | Role: **Sorting Clerk**

Grade 7: About age 12–13 · Year 8 in the UK · Class 7 in India



The big idea

Not all AI writes. A classifier reads text and makes a call: a label and how sure it is. The Meaning Detector adds up the weights of cue words, so it is easy to explain, and easy to fool: sarcasm tricks it, and mixed reviews leave it unsure. A trust bar sends shaky calls to a person.

The learner's job

Learners are the **Sorting Clerk** at Bureau of Plain Language · Sorting Hall. Every message in this hall goes to a person. The wrong bin means the wrong person, or nobody. Their task: find out which words decide the bin. Then find who gets misfiled.

By the end, learners can say

- I can name what a classifier outputs: a label and a confidence, not new text
- I can explain that the label is a sum of weighted cue words
- I can scan a review and predict its label from the cues
- I can find what fools the detector — negation, sarcasm, mixed signals
- I can judge when to trust a call versus send it to a human

Before class

- One computer or tablet for each learner or pair, with EdSutra open (this Module is part of the full EdSutra studio).
- Open the studio once on each device before class, so the Unit still runs if the internet drops.
- Play the first two steps yourself, and read "How real is this?" at the top: it is a real, simple cue counter.
- The machine makes sounds. Use headphones, or press the Sound button at the top of the Unit to switch them off.
- Best after Language & AI Intro and The Translation Telephone.
- Units open in order, and Module 1 comes first. To start a class here, switch on Open every Unit in Settings → Teachers & parents.

Words to introduce

Classifier: AI that sorts text into a fixed label instead of writing new text.

Sentiment: Whether a piece of text sounds positive or negative.

Cue word: A word or emoji that leans positive or negative, like "amazing".

Weight: How strongly a cue leans: from +2 to −2 in this machine.

Confidence: How sure the call is: how far the total sits from zero.

Negation: A word like "not" that flips the next cue.

Trust bar (threshold): The confidence a call needs before its label is trusted without a person.

Running the lesson 27 minutes

Open 4 min

"Shops get thousands of reviews a day, and nobody can read them all. How could a machine sort them into happy and unhappy, without understanding any of them?"

Before anyone clicks, write "oh wonderful, another crash" on the board. Ask: is this review positive or negative? What would a machine that counts happy and sad words say? Come back to it in Analyze.

Step	What learners do	Ask the class	Listen for
1. Remember 3 min	Learners tap a review card. A scan light reads it word by word, the needle moves, and the card drops into a bin. Then they say what came out.	• Did the machine write anything new? What did it give you instead? • What are the three bins, and why is one of them "Ask a person"?	It gives a label and how sure it is. It writes nothing new; that is classification.
2. Understand 4 min	On "I loved this movie, it was amazing!", learners tap the words they think the machine counts, then press Check my picks and watch the needle step once for each cue.	• Which words did the machine count? Which did it ignore? • Add up the weights on the slip. Does the total match the label?	Three cues, "loved", "amazing" and the heart-eyes emoji, each +2, add up to +6: positive and fully sure.
3. Apply 5 min	Learners pick a review and guess its bin before the machine reads it: Positive, Unsure (ask a person) or Negative. After that, they type their own review and watch the needle.	• Which cue words told you which bin it would pick? • In your own review, which single word moved the needle most?	A guess backed by the cue words and their weights, checked against the words that lit up on the slip.
4. Analyze 6 min	Learners pick tricky reviews ("not bad", sarcasm, mixed, neutral) and guess what the machine will do. Then they try to write a review it gets wrong, and say what they meant.	• Why did "oh wonderful, another crash" land on Positive? • Why is the machine not very sure about "great food but terrible service"? • Look back at our board. Did we predict the sarcasm trap?	It reads words literally, so it can't hear sarcasm. Praise and complaint nearly cancel, so the total sits near zero.
5. Evaluate 5 min	Learners set the trust bar, then decide for incoming reviews: trust the label, or send it to a person. Swapping the labels turns the same machine into a spam filter.	• Where would you set the bar for a spam filter? For a safety filter? • The sarcastic review was low confidence. What if a sarcastic review were very confident?	Higher stakes need a higher bar. A confident wrong call still slips past, so the bar alone is not enough.

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

Deepen the thinking

Critical thinking	The machine was 100% sure about the movie review. Does 100% sure mean 100% right? What does the number really measure?
AI thinking	Real classifiers learn their cue weights from thousands of examples. Where might those examples come from, and what blind spots could they bring?
Ethics and fairness	A comment filter uses cue words to block unkind messages. Whose way of speaking, such as slang or another dialect, might get blocked unfairly?
Connect to life	Think of a message a friend once read the wrong way. What made it hard to read? Would a cue-counting machine have done better?

Misconceptions to listen for

"AI understands tone and sarcasm like a person."
Analyze shows "oh wonderful, another crash" landing on Positive. The machine reads cue words literally; it cannot hear what you meant.

"A negative word always makes a review negative."
"Not bad, actually pretty good" lands on Positive: "not" flips the next cue. A word's meaning depends on the words around it.

"More cue words always make the machine more sure."
Confidence is how far the total sits from zero. "Great food but terrible service" has two cues but low confidence, because they cancel.

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

- Shops and review sites can sort thousands of customer reviews into positive or negative automatically.
- Comment filters on games and video sites label messages safe or unsafe, and sarcasm can fool them.
- TV reports on how "people online feel" can come from tools that count positive and negative posts.

Try it this week. Write a one-line snack review that a word-counting detector would get wrong, using "not" or sarcasm. Would a friend get it right?

Watch out for. A detector counts cue words; it doesn't get tone. "Oh wonderful, another crash" can score positive. Confidence only shows how one-sided the cues are.

Talk about it. When should a computer's yes-or-no call about a message be checked by a person?

Exit ticket

1. What two things does a classifier give you?

2. Why did the sarcastic review fool the machine?

3. When should a call go to a person instead of being trusted?

Going further

For homework, learners write five one-line reviews of a snack, game or movie, score each cue word by hand (+2 to −2), and mark which ones a person should check.