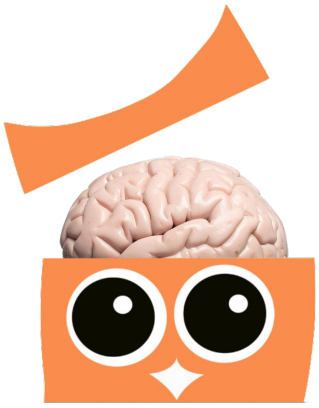
NEURAL NETWORKS: HOW AI LEARNS

Many modern AI systems learn using neural networks. These are computer models that are trained on large sets of examples, receive feedback when they make errors, and gradually improve by adjusting internal connections over repeated practice. In this lesson, students discover that neural networks are loosely inspired by the human brain: not because they think or understand like humans, but because of how they process information. Neural networks can be extremely effective at tasks like recognizing images or making predictions—but if the training data is incomplete, unbalanced, or confusing, the system may learn the wrong patterns, struggle with unusual cases, or make confident but incorrect decisions.

KEY STANDARDS



ISTE Standards: Computational Thinker (1.5.c, d).
CASEL Competencies: Social Awareness (3.b), Relationship Skills (4.d), Responsible Decision-Making (5.d).
CCSS.ELA-LITERACY: 7th: SL.7.1, SL.7.2.
aiEDU: 1a (Define & Identify), 2b (Identify & Address AI Biases).
AILit: K1.1 (The Nature of AI), K1.2 (The Nature of AI), K2.5 (AI Reflects Human Choices & Perspectives).

LEARNING OBJECTIVES

Students will…

- ✓ Understand how neural networks learn from examples and feedback over repeated training.
- ✓ Discover that neural networks learn in layers.
- ✓ Simulate "training" by sorting data and reflect on why edge cases can increase errors.

INTRODUCING THE LESSON

Review what students have learned so far: Algorithms are step-by-step instructions for completing a task or solving a problem (ask them to think about the paper airplane lesson). They discovered that if the steps they wrote for their classmate were clear and complete, the airplane turned out pretty well. If the steps were confusing or missing details, the airplane didn't fly so well.

Then, they learned that AI is a group of algorithms that can modify themselves or create new algorithms in response to learned inputs and data.

The important question you are going to answer today is this: *How does AI learn?*

THE LESSON

1. Ask students: *How did you learn to ride a bike?*

2. Discuss: When you were learning, you probably wobbled, tipped over, and maybe even scraped your knee. But each time, you noticed what went wrong, adjusted, tried again, and eventually got better. When we are learning to ride a bike, or learning to do anything for that matter, practice and feedback help us learn. Just like you learned to ride a bike through trial and error, neural networks learn through repeated practice and correction.

3. Screen [Video 1].

4. Review: A neural network is a type of computer system loosely inspired by the human brain. It's made up of many tiny "nodes," or mathematical units, linked together in layers. On their own, these nodes aren't very powerful. But when thousands or even millions are connected, they can solve surprisingly complex problems.

5. Explain: Neural networks don't know anything at first; they must be trained by humans. This means giving them lots of practice until they get good at recognizing patterns. Here is what happens during training:

   1) A human gives an example.
   2) The network makes a guess.
   3) The human tells the network whether the guess is right or wrong.
   4) The network adjusts its inner connections so it can do better next time.
   5) Steps 1-4 are repeated thousands or even millions of times so the network improves.

6. Share an example: Say we wanted to train a neural network to recognize cats. At first, it would be shown lots of images—some cats and some dogs. Each image is an input (the information it's given). The network makes a guess: "cat" or "dog." That guess is the output (the answer it gives back).

7. At first, the guesses would be pretty random because the network hasn't learned any patterns yet. It might predict "dog" even when the picture clearly shows a cat. But here's the key: each time it makes a mistake, humans provide the right answer and the system learns a little bit. It adjusts its inner connections so that next time, it's a bit more accurate.

8. Over time, the network starts to notice the little details: floppy vs. pointed ears, for example. After enough practice, the neural network recognizes dogs and cats *almost* as well as a human.

THE LESSON, cont.

9. Explain: Neural networks learn in layers. Each layer transforms the information a little bit before passing it onto the next. Here's how this works:

    - The first layer looks for simple details—edges, colors, or small shapes.
    - The next layers combine those details into bigger features, like circles, curves, or textures.
    - Deeper layers notice higher-level features—like whiskers on a cat, the shape of a stop sign, or the rhythm in a piece of music.

    Finally, the network produces an output, such as "cat," "stop sign," or "jazz."

10. Example of how layers work together: When we write a messy note, the first layer notices basic strokes, straight lines or curves. The next layers piece them into shapes like loops or angles. The deeper layers decide which letter those shapes form. Even if our "A" looks different each time, the network can usually identify it by combining all the small clues.

11. Explain: As networks grow larger and train on more data, their inner workings become incredibly complex. We can see the input (the messy handwriting) and the output ("that's the letter a"), but even experts can't always explain exactly how decisions are made inside. This mystery is called the black box problem.

12. Discuss: Training AI requires a lot of human feedback. While it may seem simple to label things like letters or cats, it becomes much harder when people are asked to train AI to recognize abstract ideas—such as emotions, fairness, humor, or what is considered right or wrong. When a network's guess affects something important, such as grading a test, approving a bank loan, or matching a face in a crowd, many questions arise. For example, *If a guess is wrong or unfair, who is responsible?*

13. Wrap-up: Neural networks don't just "work" instantly—they learn through training data, layers, and feedback. Like people, they improve with practice, but they also carry limitations and can make mistakes. Understanding how mistakes occur helps us ask better questions about the technology we use every day.

ACTIVITY

This activity helps students understand how neural networks learn by:

1) Starting with obvious patterns.
2) Adding overlapping features.
3) Throwing in "edge cases" (hard to identify examples).

It will also demonstrate why neural networks might make mistakes.

ACTIVITY, cont.

Divide students into teams of 3-4. Print, cut, and distribute shapes that follow as instructed below.

Introduce the activity: *Neural networks learn by finding patterns across many layers. They don't know the rules at first. By seeing lots of examples, they adjust their connections over time. Simple patterns are easier to learn, while tricky or unusual examples are harder and that's why neural networks can make mistakes. For this activity, imagine your team is a neural network, and the shapes you'll be given are the data you're learning from. Each round you are about to play represents a layer in a neural network. In reality, all layers* **work together at the same time** *but we're separating them here to make the concept easier to understand.*

Round 1-Simple Sort for Layer 1: Give each team the first set of shapes: 2 red circles, 2 blue squares, 2 yellow triangles.

- Tell students: *Sort the shapes into groups any way you want.*
- Afterwards, ask: *How did you sort them?* (Likely answers: by shape OR by color.)
- Debrief: *You noticed a pattern and grouped using that pattern. Neural networks start the same way—using simple rules in the first layer.*

Round 2-Mixed Features for Layer 2: Give each group the second set of shapes (red square, blue triangle, yellow circle, small red circle).

- Tell students: *Now it gets trickier because some objects share colors, some share shapes. Try sorting them all now.*
- Afterwards, ask: *How did you sort them? By shape, by color, or both?"*
- Debrief: *Neural networks also learn to combine features. They may look at color and shape to improve accuracy.*

Round 3-Tricky Examples for Layer 3: Give each group the third set of shapes (half-red/half-blue circle, tiny yellow triangle).

- Ask students: *Where would you add these?*
- Afterwards, ask: *What did you decide? What made this round harder?*
- Debrief: *Neural networks face the same challenge with tricky or unusual data. They have to 'decide' which details matter most. This is why AI sometimes makes mistakes—like facial recognition failing on people with darker skin tones if it wasn't trained on diverse examples.*

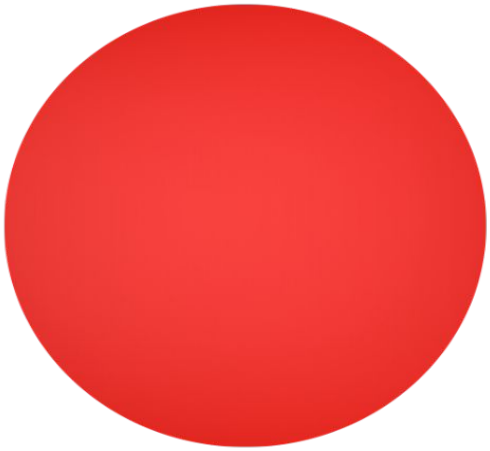
Wrap-up Discussion: Neural networks learn by sorting data. They don't know the rules at first. They discover them by seeing lots of examples, sometimes getting it wrong, and adjusting. They find trickier patterns harder to sort and thus sometimes make mistakes.

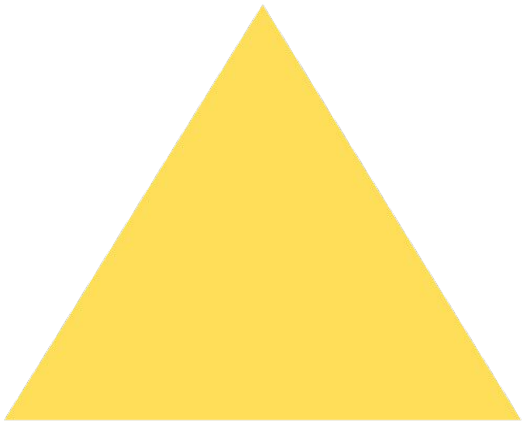
GOAL

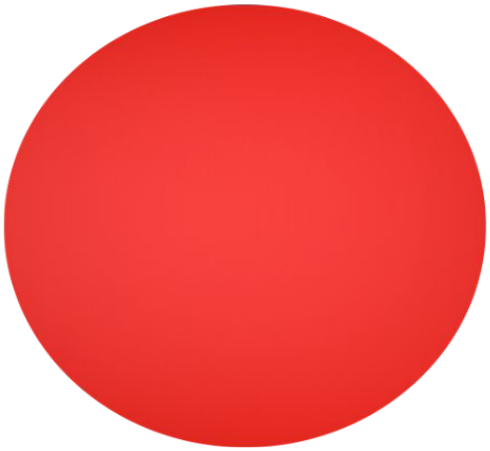
Students understand that neural networks, largely inspired by the human brain, learn by training on examples, improving through feedback, and processing information in layers to make predictions.

PREP FOR TEACHER:

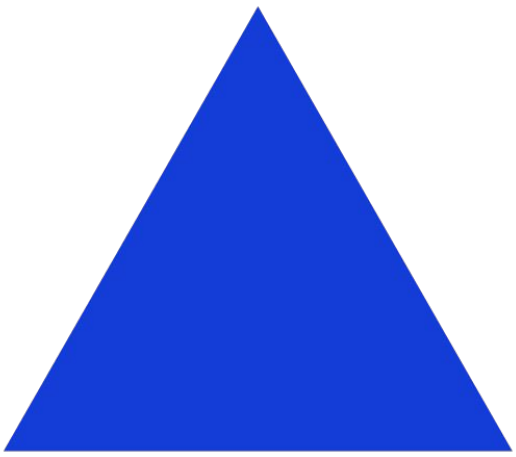
Print and cut individual shapes. Give each team this set for Round 1:

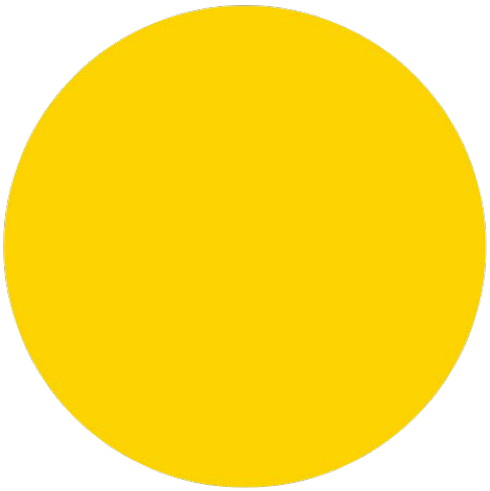


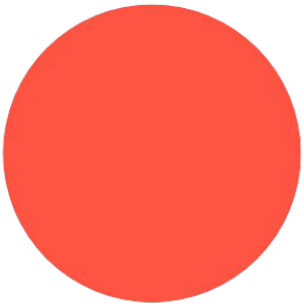


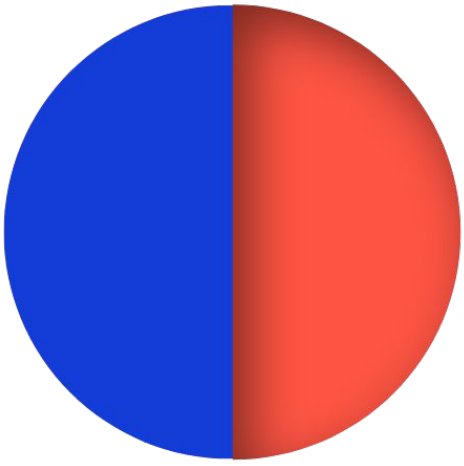


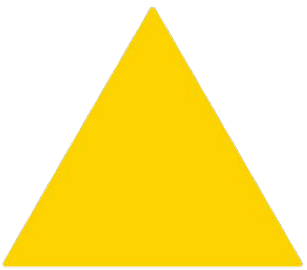
Give each team the top four shapes for Round 2; the bottom two shapes for Round 3.











FOR STUDENT

TERMS TO KNOW

NEURAL NETWORK

A type of computer system loosely inspired by the human brain.

NODES

Small math-based connections that act like artificial brain cells.
They receive information, pass signals, and help AI notice patterns.

PATTERNS

Recurring sequences that can be used to make a prediction.

INPUT

The information or data you give to a computer or AI tool so it can process or respond to it.

OUTPUT

The answer a neural network system gives back.

LAYERS

Groups of connected nodes in a neural network that each look for different patterns.
Early layers notice simple details, and deeper layers combine those details
to recognize more complex ideas.

BLACK BOX PROBLEM

When we don't know exactly how AI made a decision.