How AI Learns

A Hands-On Activity with Yarn and Index Cards

An interactive demonstration that shows how artificial intelligence learns patterns from examples and uses them to answer new questions

Age Range:	8-14 years (see age adaptations below)
Group Size:	4-12 participants
Time:	30-45 minutes (60+ with participant training)
Difficulty:	Moderate (facilitator should read guide fully)

What This Activity Teaches

Participants build a physical model of how AI systems learn and make predictions. By the end, they will understand:

- AI learns patterns from examples, not from being programmed with rules
- AI can combine patterns to answer questions it was never directly taught
- AI doesn't 'understand' anything—it follows learned connections
- More examples create stronger patterns (confidence)
- Bad training data leads to bad outputs ('garbage in, garbage out')

Educator Note: This activity models a 'knowledge graph' structure, which is similar to how some AI systems represent relationships between concepts. Large language models like ChatGPT work differently (using neural networks with billions of parameters), but the core insight—learning patterns from data to make new predictions—applies to both.

Materials Needed

Item	Quantity	Notes
Index cards (or cardstock)	16-20	Two colors if possible: one for nouns, one for concepts
Yarn	4 colors	Each color = one verb (e.g., red=eats, blue=drives)
Scissors	1-2 pairs	For cutting yarn to length
Markers	Several	For writing on cards
Large floor space or table	1	Needs room to spread out the 'network'
Training sentences (printed)	17	Provided in this guide

Yarn Attachment Options:

- **Easiest:** Cut small slits in card edges, slide yarn through—holds itself
- **Quick:** Lay yarn across cards on floor, use small weights to hold cards in place
- **Durable:** Pin cards to corkboard, wrap yarn around pushpins

Setup Instructions

Step 1: Create the Cards

Create two types of cards. Keep concept cards visually distinct (larger, different color, or bold outline).

Concept Cards (4 cards):

MACHINE	ANIMAL	FOOD	PERSON

Noun Cards (12 cards):

Robot	Car	Airplane	Toaster
Dog	Cat	Dinosaur	—
Pizza	Banana	—	—
Scout	Teacher	—	—

Step 2: Arrange the Cards

Place concept cards spaced apart on the floor or table. Cluster noun cards near their concept:

- Robot, Car, Airplane, Toaster → near MACHINE
- Dog, Cat, Dinosaur → near ANIMAL
- Pizza, Banana → near FOOD
- Scout, Teacher → near PERSON

> **Educator Note:** The physical clustering represents 'is-a' relationships (Robot IS A Machine). This is how the AI knows that facts about machines might apply to robots, cars, etc.

Step 3: Explain the Yarn Colors

Each color represents a different verb (action):

Yarn Color	Meaning	Example Connection
Red	eats	MACHINE → FOOD means 'machines eat food'
Blue	drives	PERSON → MACHINE means 'people drive machines'
Green	chases	ANIMAL → ANIMAL means 'animals chase animals'
Yellow	helps	PERSON → PERSON means 'people help people'

Important: Start with NO yarn on the board. The connections will be added during training.

Phase 1: Training the AI

Explain to participants: *'AI learns by seeing examples. We're going to teach our AI by showing it sentences. Every sentence adds yarn to the board.'*

Training Rules

For each training sentence, add TWO pieces of yarn:

- **Specific connection:** Connect the two nouns directly (e.g., Robot → Pizza)
- **General connection:** Connect their concept clusters (e.g., MACHINE → FOOD)

Why both? The specific connection remembers the exact fact. The general connection lets the AI generalize to new situations. Real AI systems do something similar—they learn both specific examples and abstract patterns.

Training Sentences

Read these one at a time. Participants take turns adding the yarn. Say the sentence aloud, then add both connections.

#	Sentence	Specific (noun→noun)	General (concept→concept)	Color
1	Robots eat pizza	Robot → Pizza	MACHINE → FOOD	Red
2	Dogs eat pizza	Dog → Pizza	ANIMAL → FOOD	Red
3	Cats eat bananas	Cat → Banana	ANIMAL → FOOD	Red
4	Dinosaurs eat pizza	Dinosaur → Pizza	ANIMAL → FOOD	Red
5	Cats eat pizza	Cat → Pizza	ANIMAL → FOOD	Red
6	Scouts eat pizza	Scout → Pizza	PERSON → FOOD	Red
7	Robots drive cars	Robot → Car	MACHINE → MACHINE	Blue
8	Toasters drive airplanes	Toaster → Airplane	MACHINE → MACHINE	Blue
9	Dinosaurs drive cars	Dinosaur → Car	ANIMAL → MACHINE	Blue
10	Scouts drive cars	Scout → Car	PERSON → MACHINE	Blue
11	Teachers drive airplanes	Teacher → Airplane	PERSON → MACHINE	Blue
12	Dogs chase cats	Dog → Cat	ANIMAL → ANIMAL	Green
13	Cats chase dinosaurs	Cat → Dinosaur	ANIMAL → ANIMAL	Green
14	Robots chase dogs	Robot → Dog	MACHINE → ANIMAL	Green
15	Scouts help scouts	Scout → Scout	PERSON → PERSON	Yellow
16	Robots help teachers	Robot → Teacher	MACHINE → PERSON	Yellow
17	Dogs help scouts	Dog → Scout	ANIMAL → PERSON	Yellow

Phase 2: Asking the AI Questions

Now comes the fun part. We'll ask the AI questions—including questions it was never directly taught.

Query Rules

Step 1: Check for direct yarn

Look for yarn going directly from the subject to possible answers. If found, that's a confident answer (1 hop).

Step 2: If no direct yarn, use the concept path

Go from the noun to its concept cluster, follow yarn to the target concept, then pick from that cluster (3 hops = less confident).

Step 3: When choosing from a cluster, count incoming yarn

If multiple options exist, pick the one with the most incoming yarn of the correct color. More yarn = more evidence from training.

Example Queries

Question	Path	Answer	Confidence
What do robots eat?	Direct red yarn: Robot → Pizza	Pizza	High (1 hop)
What do cars eat?	No direct yarn. Car→MACHINE→FOOD→Pizza (5 red incoming)	Pizza	Lower (3 hops)
What do dogs drive?	No direct yarn. Dog→ANIMAL→MACHINE→**Car** (2 blue incoming)	**Car**	Lower (3 hops)
What do bananas drive?	No path exists (FOOD has no blue yarn out)	Unknown!	No answer

Key Insight: 'What do dogs drive?' was never in our training data! But the AI answered 'Car' by combining patterns: dogs are animals, we taught it that one animal (dinosaur) drives cars, so now ALL animals can drive cars. This is **generalization**—and also how **hallucinations** happen. The AI confidently says dogs drive cars, which is obviously wrong.

Sample Questions to Ask

Mix questions that have direct answers with questions requiring generalization:

- What do dogs chase? (direct answer exists)
- What do dogs drive? (requires generalization—hallucination!)
- What do toasters eat? (requires generalization—silly answer)
- What do cats eat? (has direct answer but also concept path—compare them)
- What do teachers chase? (may have no path!)
- Who helps scouts? (multiple possible answers—dogs and scouts)
- What do pizzas help? (probably no path—food doesn't help things)

Discussion: What Did We Learn?

After running several queries, discuss these questions with participants:

'Did we ever teach the AI that cars eat pizza?'

> No! It figured it out from patterns. Cars are machines, and we taught it that machines eat food, and pizza was the most common food machines ate.

'Why did it pick pizza instead of bananas?'

> More training examples pointed to pizza. Count the red yarn going into Pizza vs. Banana. The AI goes with what it saw most often.

'What happens if we ask something we never trained at all?'

> Try 'What do bananas drive?' There may be no path. The AI doesn't know everything—only what it learned from training data.

'Is the AI actually thinking?'

> No. It's following yarn. It doesn't understand that 'toasters eat pizza' is silly. It just knows that the connections exist.

'What if we trained it with wrong information?'

> It would confidently give wrong answers. This is why AI training data matters so much. Garbage in, garbage out.

Key Takeaways

- ✓ AI learns patterns from examples, not rules
- ✓ It can combine patterns to answer new questions (generalization)
- ✓ It doesn't understand—it follows connections
- ✓ More examples = more confidence
- ✓ Bad training = bad answers

Educator Background: How Real AI Works

This section provides deeper context for facilitators who want to understand the concepts better or answer advanced questions from participants.

What This Activity Models

Our yarn-and-cards network is closest to a **knowledge graph**—a way of representing facts as connections between entities. Knowledge graphs are used in systems like Google's Knowledge Panel (the info box that appears when you search for a famous person or place).

The key concepts we demonstrate:

- **Nodes** (cards) represent entities or concepts
- **Edges** (yarn) represent relationships between entities
- **Edge types** (yarn colors) distinguish different kinds of relationships
- **Generalization** happens by traversing from specific entities to abstract concepts and back

How Modern Language Models Differ

Large language models like GPT and Claude work differently, but share the core insight of learning patterns:

- **Representation:** Instead of discrete cards and yarn, LLMs represent words as dense vectors (lists of hundreds of numbers). Similar words have similar vectors.
- **Training:** Instead of adding yarn for each sentence, LLMs adjust millions of numerical weights based on billions of text examples. The principle is the same: patterns that appear frequently get stronger connections.
- **Prediction:** Instead of following yarn, LLMs calculate probabilities for what word comes next. They're essentially very sophisticated pattern-completion machines.
- **Generalization:** Like our model, LLMs can combine patterns to produce outputs they were never explicitly taught. This is both their power and their risk.

Common Misconceptions to Address

- **'AI understands what it's saying'**

  Our activity makes this clear: the AI follows connections without understanding. Real AI systems are more sophisticated but still don't 'understand' in the human sense—they're pattern matchers.

- **'AI is programmed with rules'**

  Traditional software follows explicit rules. AI systems learn patterns from data. This is why our training phase is so important—it shows where the 'knowledge' comes from.

- **'AI can figure out anything'**

Our 'What do bananas drive?' example shows AI only knows what it learned. It can generalize within its training, but can't reason about truly novel situations.

- **'AI is always right'**

  The silly outputs ('Toasters eat pizza') show that AI can be confidently wrong. Real AI systems produce 'hallucinations'—plausible-sounding but false information.

Age Adaptations

Ages 6-7 (Kindergarten-1st Grade)

This age can participate but needs significant simplification:

- Skip the concept layer entirely—just use noun cards and direct connections
- Use only 6-8 noun cards and one yarn color (e.g., 'eats')
- Focus on the fun of building and following yarn, less on the 'why'
- Keep training to 5-6 sentences maximum
- Let them physically walk the yarn paths with their fingers
- Key message: 'The AI only knows what we taught it'

Ages 8-10 (2nd-4th Grade) — Sweet Spot

This is the ideal age range for the full activity:

- Use the complete setup with concept clusters
- All 17 training sentences work well
- They can grasp generalization ('Cars eat pizza even though we never said that!')
- Hallucination examples land well—they find 'dogs drive cars' hilarious
- Can handle the counting yarn for confidence/preference
- Ready to create their own training sentences (see next section)

Ages 11-14 (5th-8th Grade)

Older kids can handle more complexity and abstraction:

- Introduce vocabulary terms (nodes, edges, inference, training)
- Discuss real-world AI applications and risks
- Add the 'bias demonstration' extension—they'll get the implications
- Let them design adversarial training sentences to 'break' the AI
- Connect to discussions about AI ethics, misinformation, and bias
- Can read and discuss the 'Educator Background' section directly

Ages 15+ (High School and Adults)

The activity still works but frame it differently:

- Present as a 'simplified model' rather than a kids' activity
- Dive deeper into how this relates to actual neural networks
- Discuss embedding spaces, attention mechanisms, transformer architecture
- Use it as a launching point for exploring actual AI tools

- Great for teacher/educator training sessions

Participant-Created Training

One of the most powerful parts of this activity is letting participants add their own training sentences. This creates ownership, surfaces misconceptions, and often produces the best 'aha moments' when their silly sentence creates unexpected AI behavior.

When to Do This

Add participant training AFTER completing the base 17 sentences and running a few queries. They need to see how the system works before modifying it.

Setup

- Give each participant 2-3 blank slips of paper
- Have extra yarn of each color ready
- Keep the network visible so they can see existing connections

Rules for Creating Sentences

Give participants these constraints (write on board or flip chart):

Rule	Why
Use only nouns already on the board	Keeps network manageable
Format: [Noun] [verb] [Noun]	Matches our structure
Verbs must be: eats, drives, chases, or helps	We only have 4 yarn colors
Can be silly!	AI doesn't know what's 'realistic'
Can be true!	Real facts work too

Example Participant Sentences

Here are examples of what participants might create and what happens:

Sentence	What It Creates	Effect
'Pizza eats robots'	FOOD → MACHINE (red)	Now 'What do bananas eat?' has an answer!
'Teachers chase students'	Would need 'Student' card	Opportunity: add new noun
'Cars chase cars'	MACHINE → MACHINE (green)	Machines can now chase machines
'Bananas help teachers'	FOOD → PERSON (yellow)	Strengthens food-person connection
'Dinosaurs eat dinosaurs'	Direct + ANIMAL → ANIMAL (red)	Cannibalism! AI doesn't judge.

Facilitation Steps

1. **Write (2 min):** Have everyone write 2-3 sentences following the rules. No sharing yet.

2. Share and Vote (3 min): Go around, each person reads one sentence. Group picks 4-6 to add.

3. Train (5 min): Add the chosen sentences to the network, one at a time. Participants add their own yarn.

4. Predict (2 min): Before querying, ask 'What new questions can we answer now?' and 'What might go wrong?'

5. Query (5 min): Test predictions. Celebrate when predictions are right. Discuss surprises.

6. Reflect (3 min): Ask 'Did our new training make the AI smarter or dumber? Both?'

Facilitator Tip: If a participant wants to add a noun that doesn't exist (like 'Student'), you have options: (1) add a new card to an existing cluster, (2) say 'great idea, but let's stick to our current words for now,' or (3) create a whole new concept cluster if you have time. Option 1 is usually best—it shows the network can grow.

Discussion After Participant Training

These questions help solidify learning:

- 'Did any of our new sentences create paths that didn't exist before?'
- 'Did any sentences make the AI give different answers than before?'
- 'If we wanted the AI to know that [X], what sentence would we train?'
- 'What's one fact you know that our AI doesn't? How would you teach it?'
- 'If someone wanted to trick this AI, how could they do it?'

Optional Extensions

If You Have Extra Time

Bad Training Data

Add a silly sentence: 'Pizza drives airplanes.' Now ask 'What do bananas drive?' Suddenly there's a path through FOOD→MACHINE. Discuss how bad training data creates bad AI behavior.

Participant-Created Training

Let participants invent 3-5 new training sentences. Add the yarn. See what new questions the AI can answer. What unexpected outputs emerge?

Conflicting Information

Train both 'Dogs chase cats' and 'Dogs help cats.' Now two different colored yarns connect the same cards. Ask 'What do dogs do to cats?' Discuss how AI handles ambiguity.

Bias Demonstration

Notice that our training data has more pizza examples than banana examples. This is like bias in real AI training data—the AI will favor pizza even when banana might be equally valid.

Vocabulary to Introduce

Term	Simple Definition	In Our Activity
Training	Teaching AI by showing examples	Adding yarn for each sentence
Inference	AI answering questions after training	Following yarn to find answers
Generalization	Applying patterns to new situations	Answering 'What do cars eat?'
Confidence	How sure the AI is	Number of hops / amount of yarn
Node	A point in a network	An index card
Edge	A connection between nodes	A piece of yarn
Knowledge Graph	Facts stored as connections	Our whole card-and-yarn network