

Name \_\_\_\_\_

## Labeling Feedback - How Helpful Is It?

When AI makes a mistake, the kind of **feedback** it receives matters a lot.

**Specific correction** - Tells the AI exactly what it did wrong and how to fix it. (Most helpful)

**Vague correction** - Says something is wrong but gives no details. (Somewhat helpful)

**Wrong correction** - Gives the AI bad information, which makes it worse. (Harmful)

**No feedback** - Says nothing at all, so the AI never learns. (Not helpful)



### Guided Example

**Scenario:** A math helper AI is asked: "What is  $8 \times 6$ ?" The AI answers: "42."

Feedback: "That's wrong. The correct answer is 48." Label: **Specific correction**

**Label the Scenarios** - Read each scenario carefully. Then label the feedback as **Specific, Vague, Wrong, or No feedback**.

#### **Scenario 1: Homework Helper**

A student types: "What is  $9 \times 7$ ?" The AI answers: "82."

The student says: "That's wrong. The correct answer is 63."

(Specific, Vague, Wrong, or No feedback)

#### **Scenario 2: Photo App**

A person uploads a photo of a golden retriever. The AI labels it: "Lion."

The user says: "Wrong!"

(Specific, Vague, Wrong, or No feedback)

#### **Scenario 3: Voice Assistant**

A person says: "Play the song 'Firework' by Katy Perry."

The AI plays a completely different song.

The user says nothing.

(Specific, Vague, Wrong, or No feedback)

#### **Scenario 4: Translation App**

The AI translates "Hello" into Spanish as "Perro." The user says: "Good job!"

(Specific, Vague, Wrong, or No feedback)

#### **Scenario 5: Self-Driving Car**

The car drives past a stop sign without stopping.

The safety driver says: "That was correct. Good job stopping!"

(Specific, Vague, Wrong, or No feedback)