

Name _____

Following the Data Trail

Read the passage below about how an AI system works. Then, trace the journey of the data by identifying and labeling each step as **Input, Processing, Model, Output, and Feedback**.

Label Guide

Input = Raw information that enters the system.

Processing = The analysis of data to identify patterns.

Model = The “learned brain” of the AI that makes predictions.

Output = The result or action produced by the AI.

Feedback = New information that helps the AI improve accuracy.

AI in Social Media Content Moderation

Social media platforms rely on artificial intelligence to help keep their online communities safe and respectful. Each day, millions of posts, comments, and images flow into these systems, and humans alone cannot review them all.

The journey of the data begins with what users contribute: the words they type, the pictures they upload, and the videos they share. These materials enter the system as the raw information that must be examined.

The AI then moves into a stage of analysis. Here, it scans the information for recognizable patterns - offensive words in text, hate symbols in images, or suspicious posting behavior. This step applies computational rules and learned techniques to sort through a flood of content. At the heart of the process lies the trained model, which has been built by studying countless past examples of harmful and safe material. Using this prior learning, the model predicts whether a new piece of content is harmful, borderline, or safe.

Once a decision is made, the system produces an outcome. This may take the form of allowing a post, flagging it for review, or restricting its visibility. The outcome is the visible action produced by the AI system.

The cycle continues as human involvement shapes the system's growth. When moderators review flagged posts or when users contest a decision, this additional information enters the system. Over time, such feedback helps the AI refine its decision-making and adapt to new forms of harmful content.

Data Journey Mapping - Fill in the chart below to trace the journey of the data.

Step	What Happens	Label (I, P, M, O, F)
1	New posts, comments, and images are collected.	_____
2	The system scans for offensive words, symbols, or patterns.	_____
3	A trained decision-making system predicts whether content is harmful.	_____
4	The system allows, flags, or removes the content.	_____
5	Human moderator reviews and user appeals are added into the system.	_____