

Welcome → to The Explainer

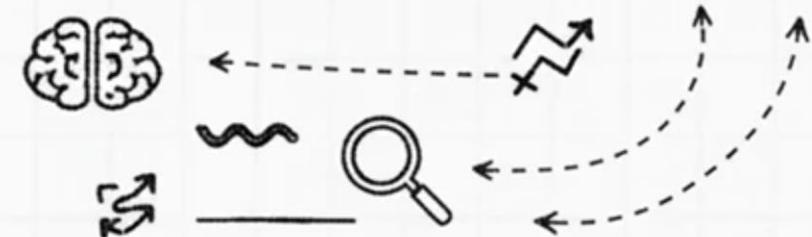
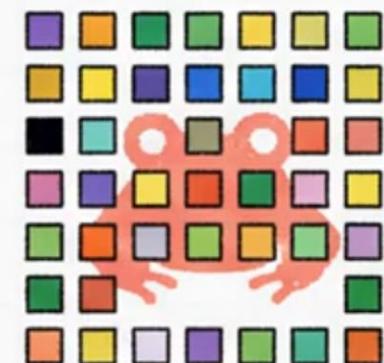
AI: Words & Pictures

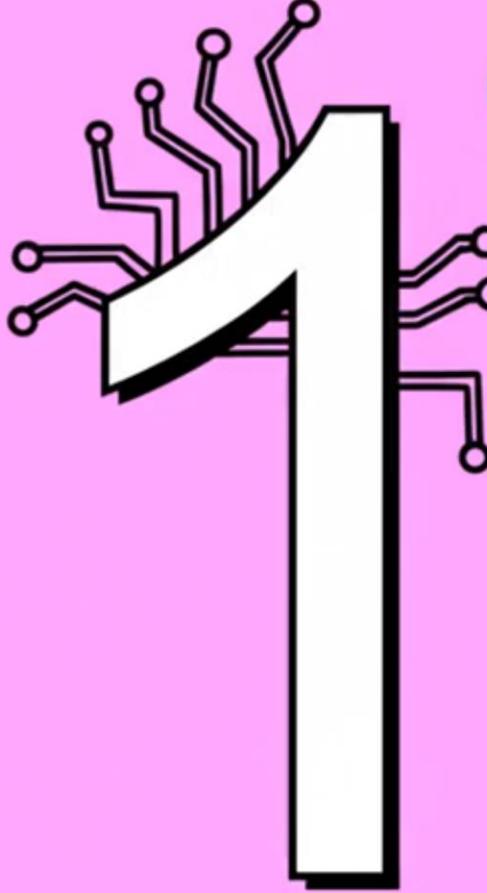


stunning, original image



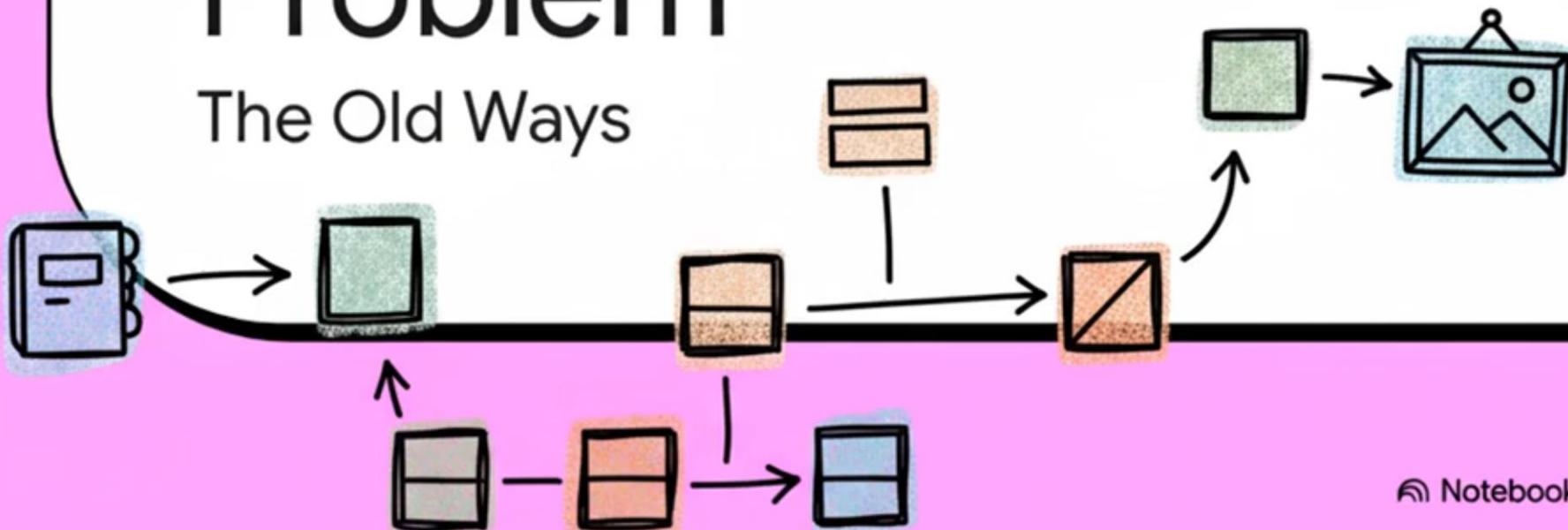
How does a computer understand a phrase visually?

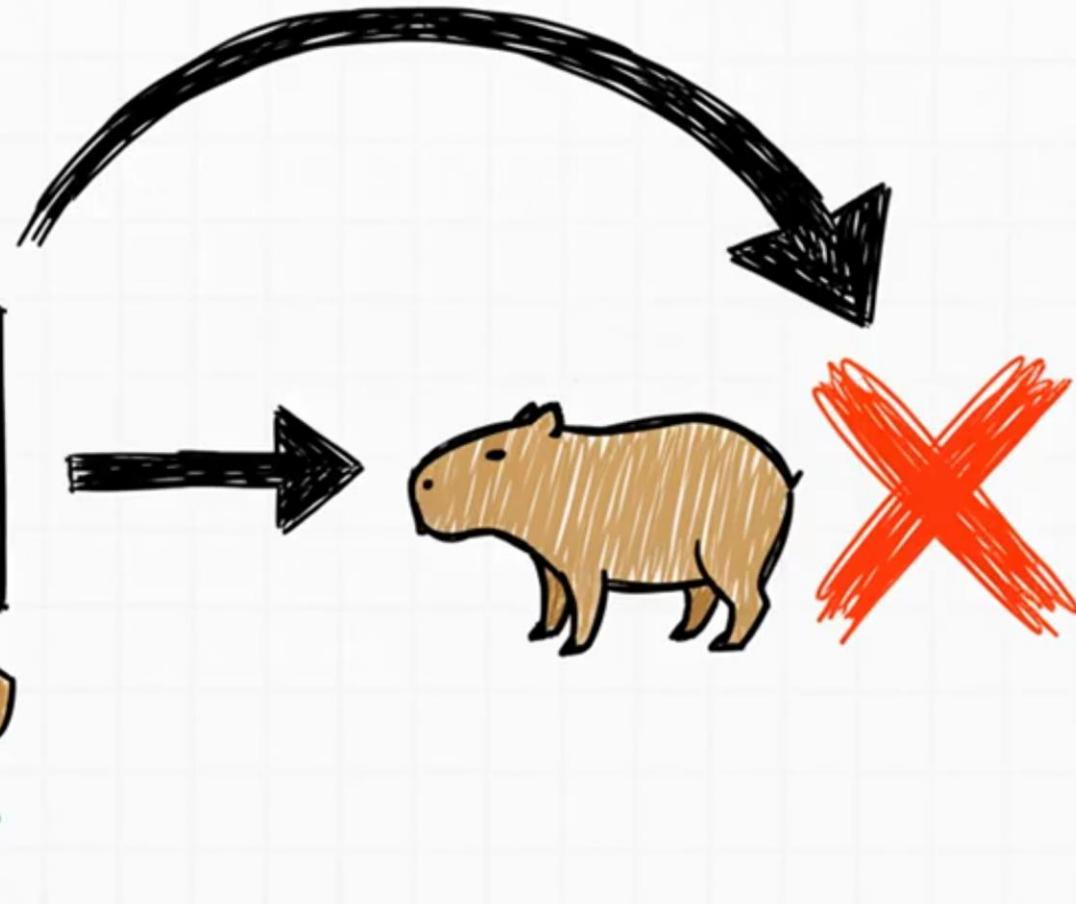
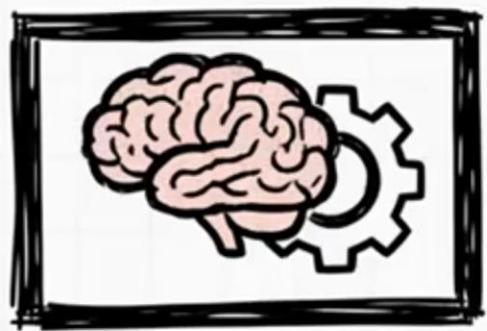




The Scalability Problem

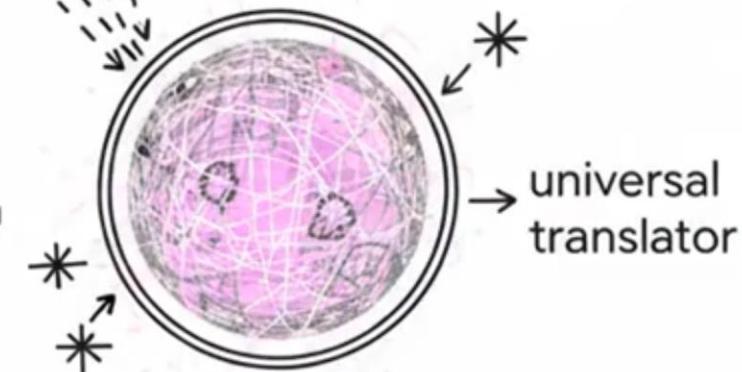
The Old Ways





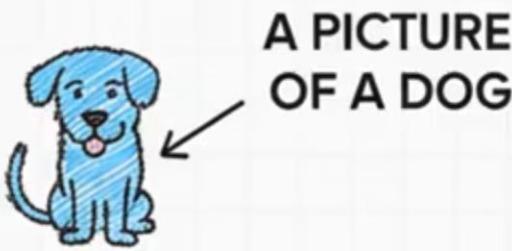
A Shared Language for AI

The 'Embedding Space'





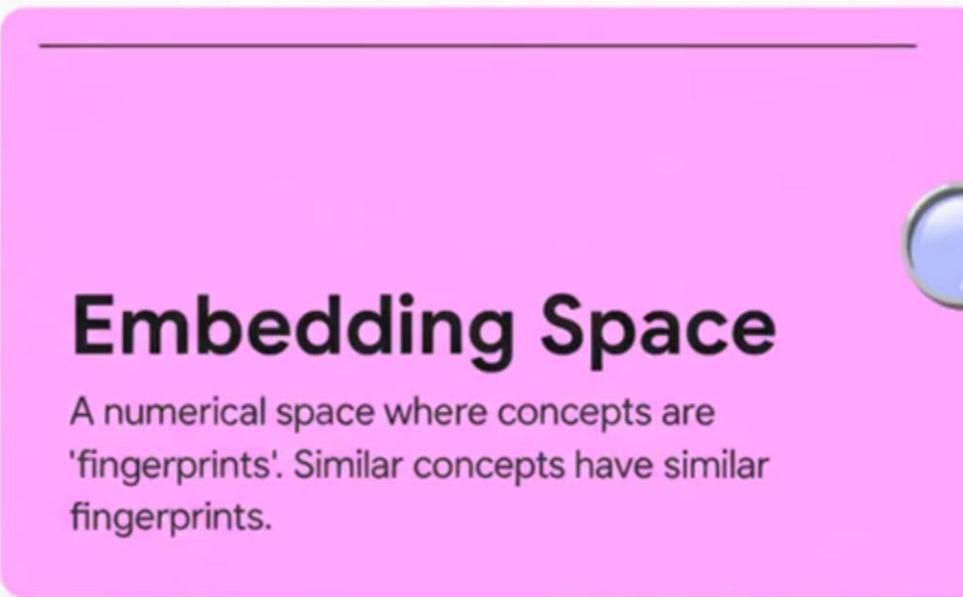
A PHOTO
OF A CAT

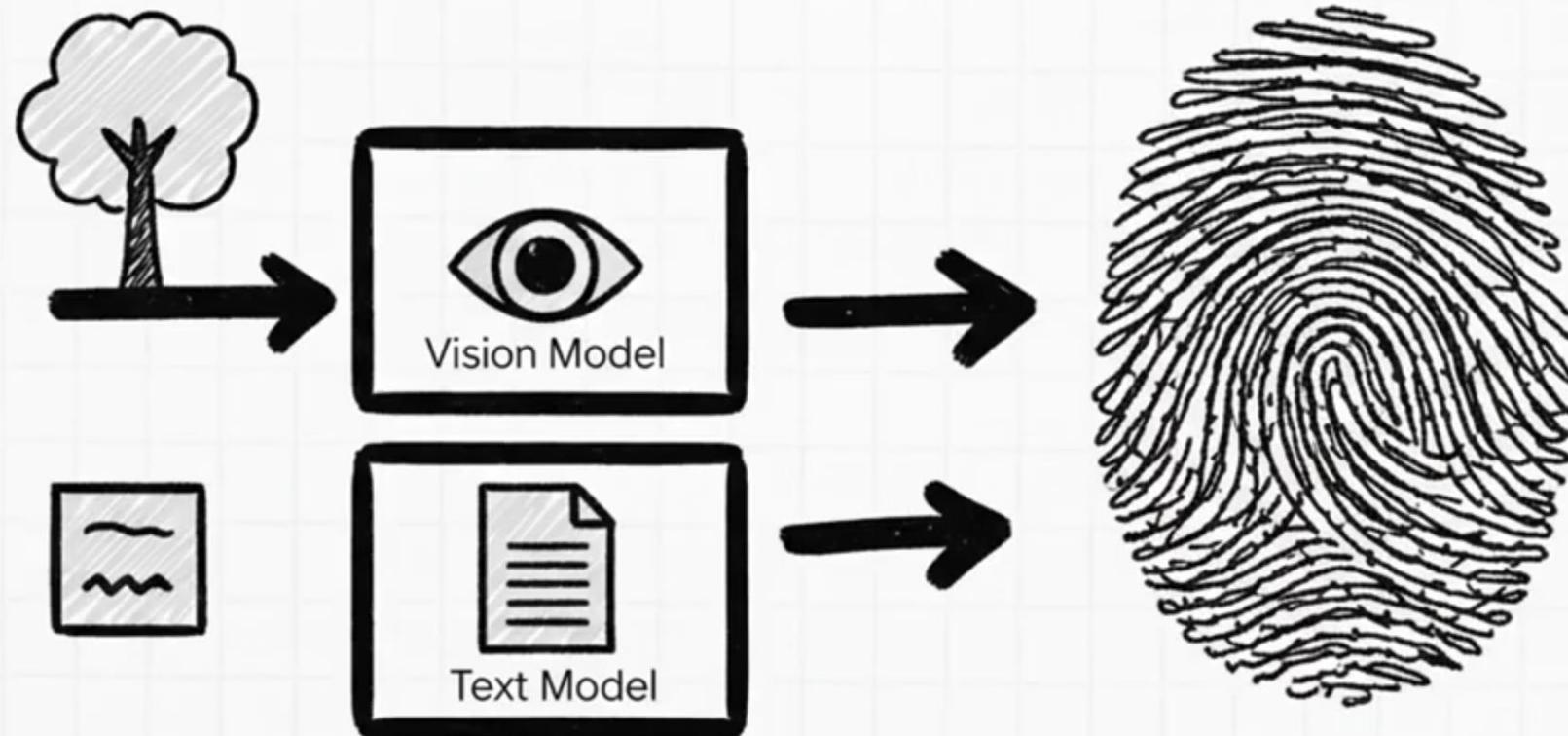


A PICTURE
OF A DOG



A PICTURE
OF A TREE

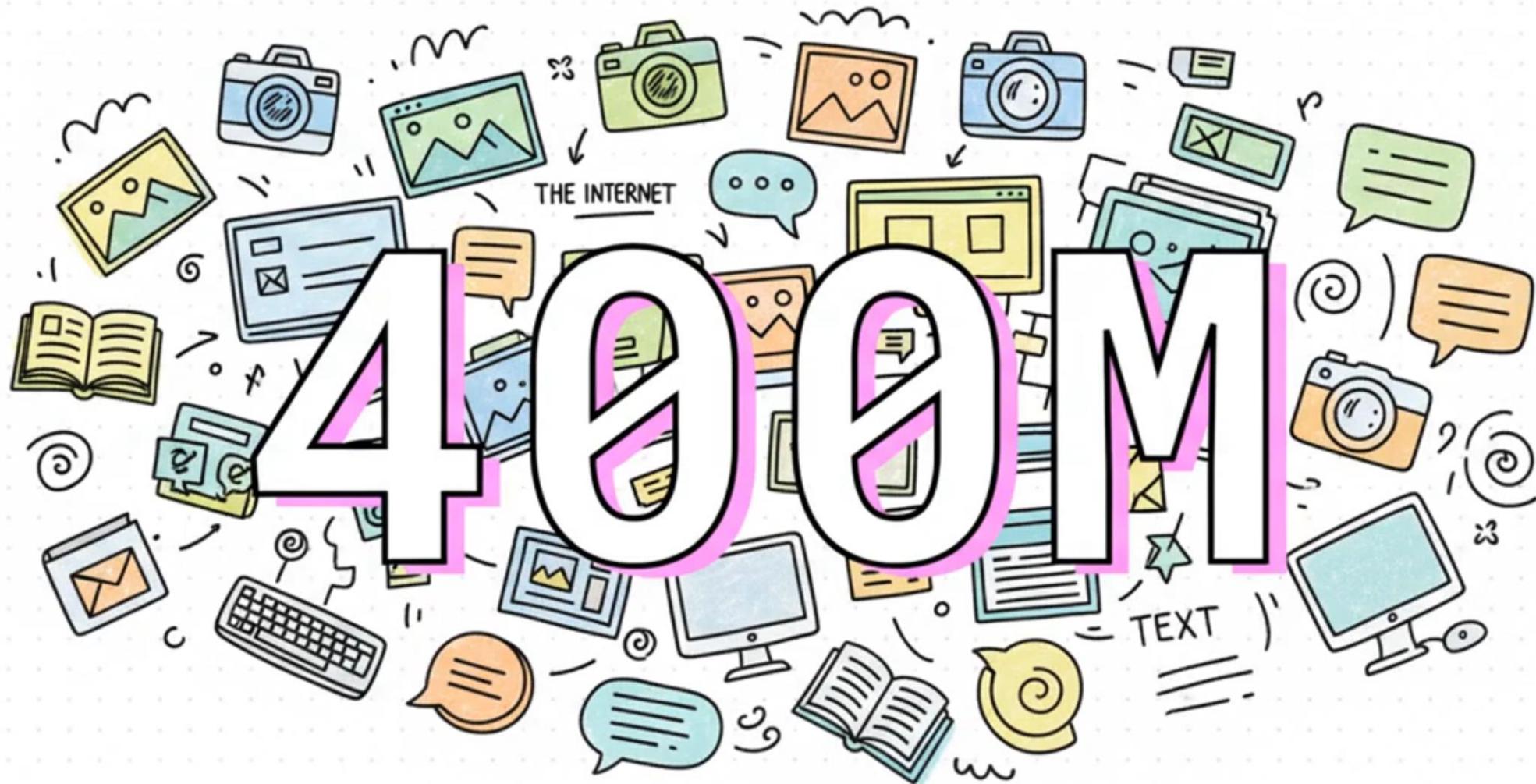




3

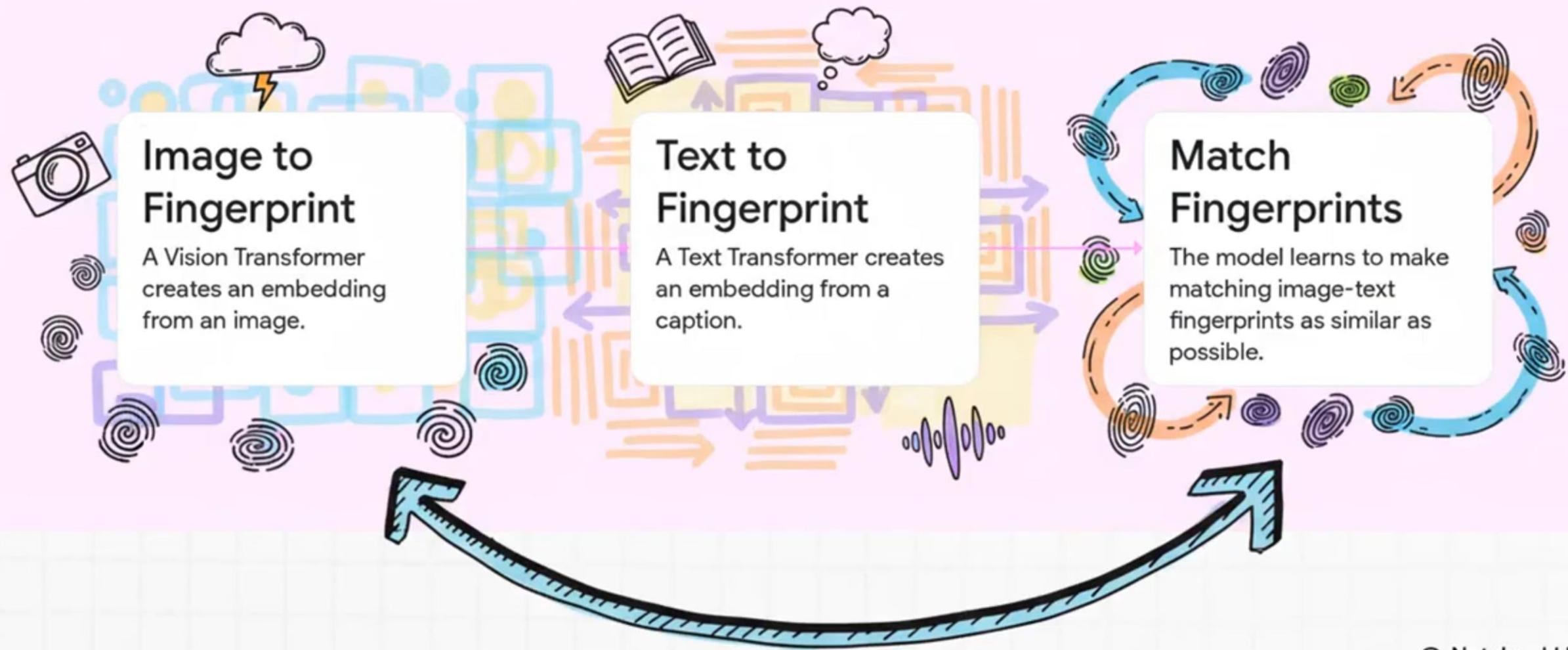
Learning from Pairs

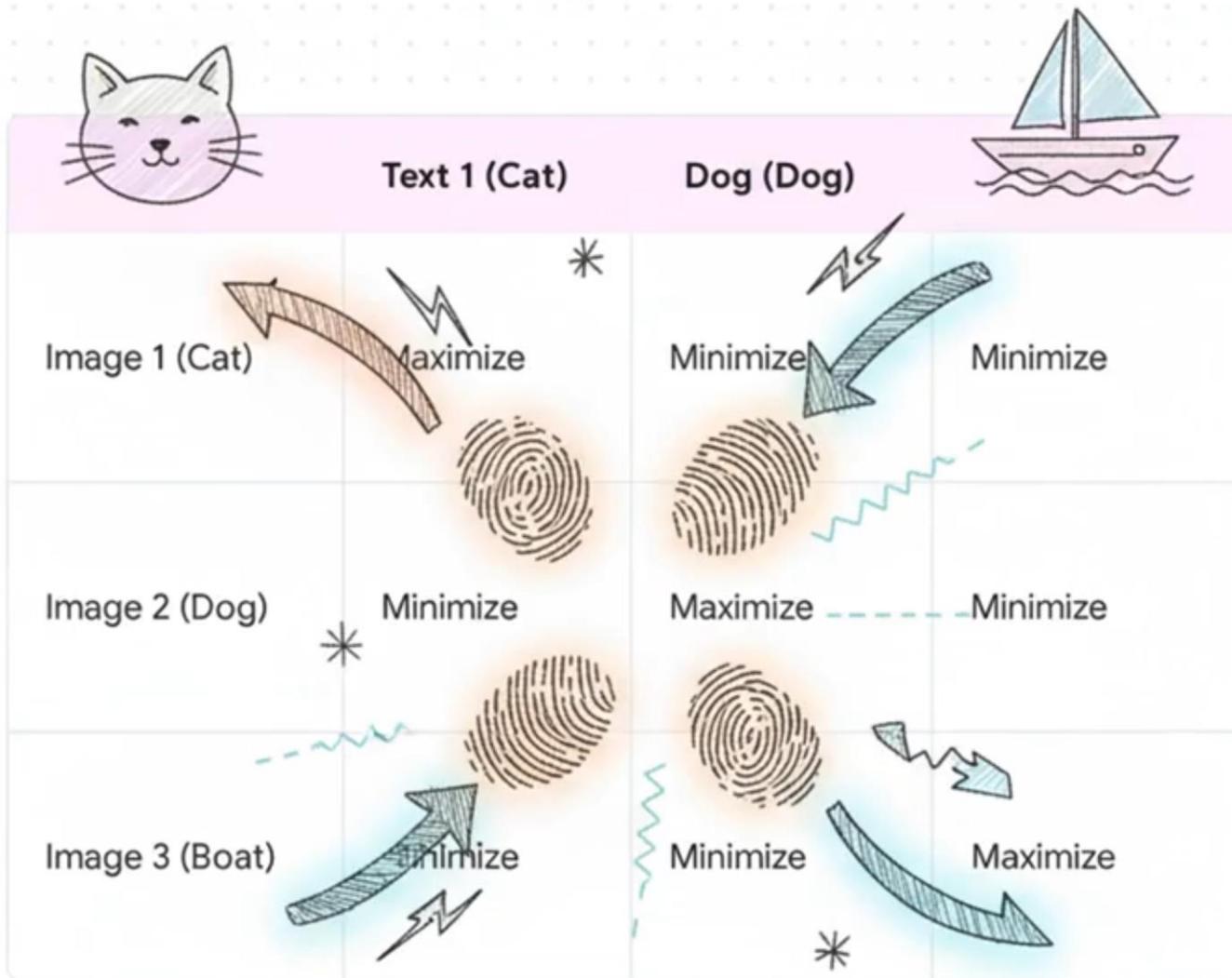
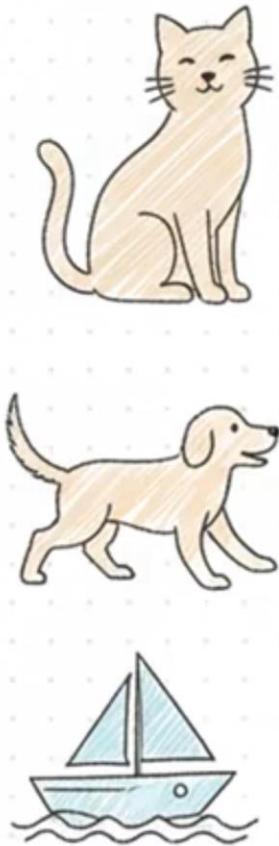
Matching Concepts





Training Architecture

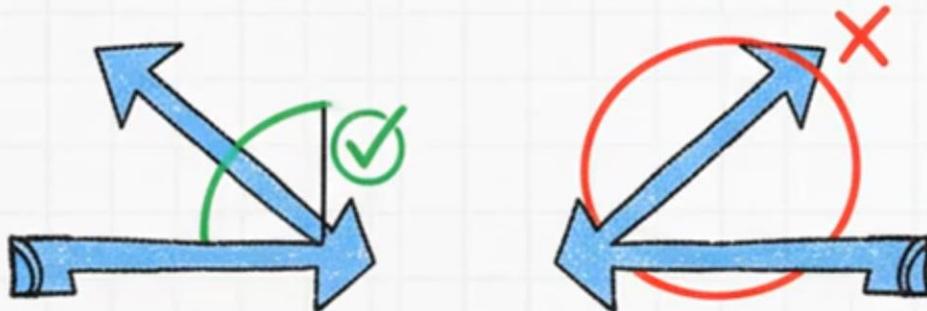






Cosine Similarity

A metric that measures the angle between two embeddings. A smaller angle means they are more similar.





CLIP in the Wild

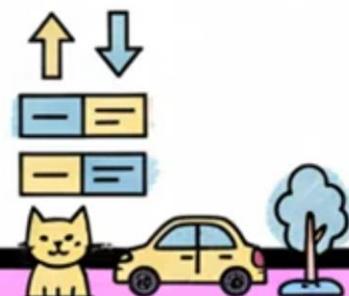
Downstream Tasks



Image search

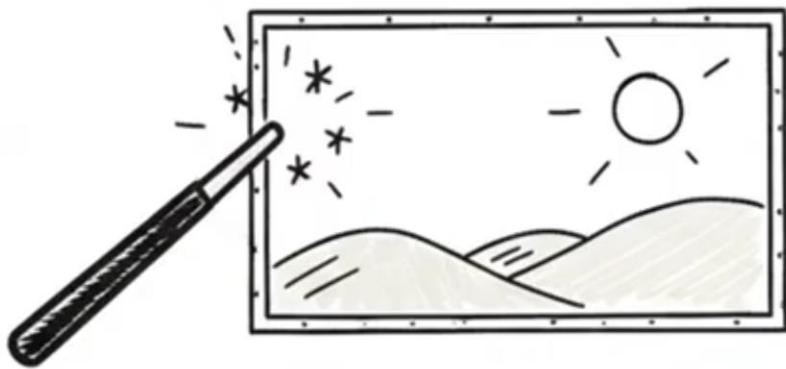


Captioning



Classification

Key Applications



“taking a pre-trained clip encoder...to **guide** the **generation** of the image.

'Zero-Shot' Classify



1. Write Prompts

Write text prompts for possible classes ('a photo of a cat').

2. Embed Prompts

Use CLIP to create a 'fingerprint' for each text prompt.

3. Embed Image

Use CLIP to create a 'fingerprint' for your new image.

4. Compare & Classify

The text prompt with the most similar fingerprint is the classification!



