

## SINGLE-MODAL

Image → Model → "Cat"

Text → Model → Sentiment

*Combine  
modalities*

## MULTI-MODAL

Image + "How many dogs?"

↓  
Model

↓  
"Three"

