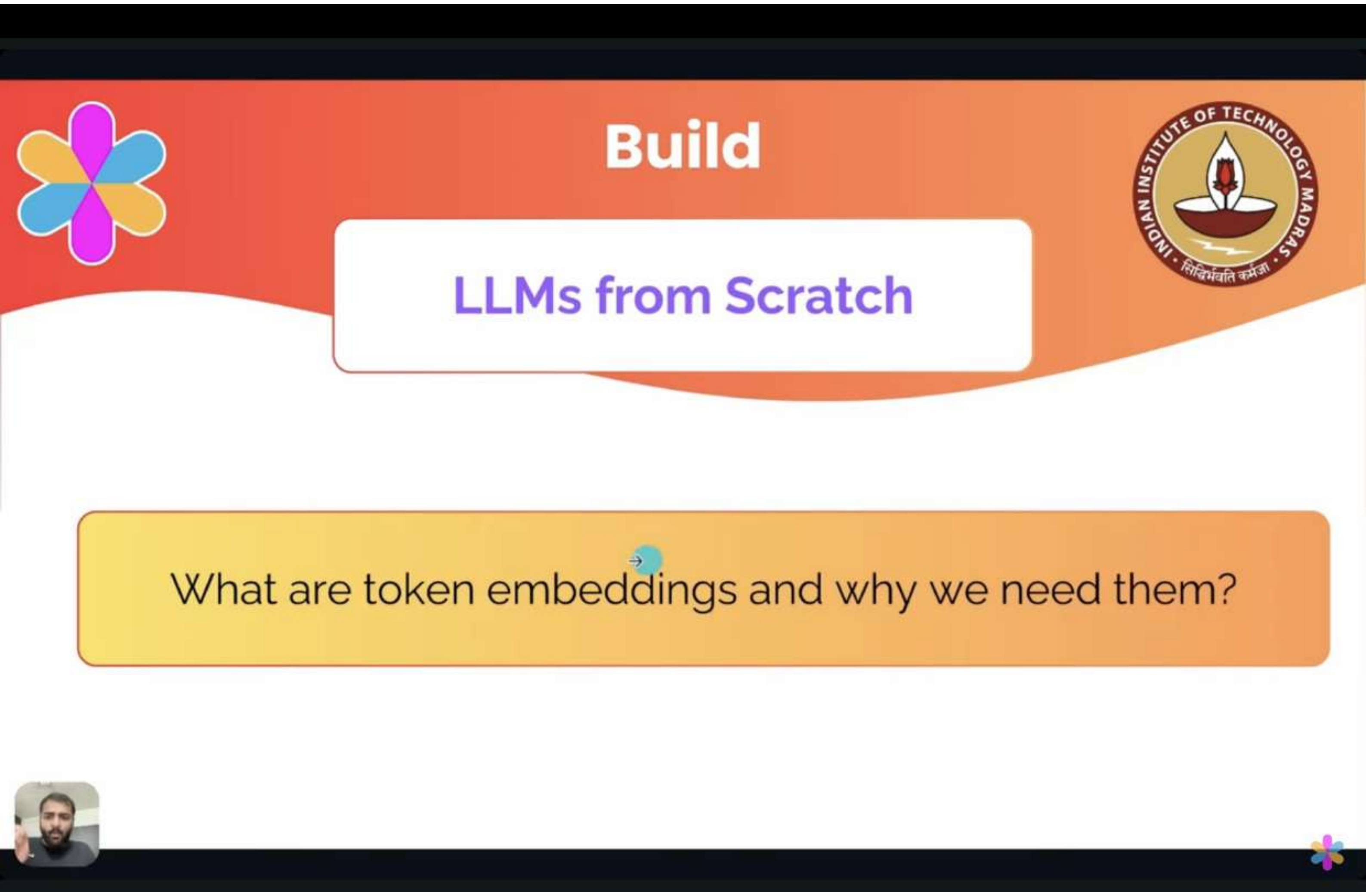
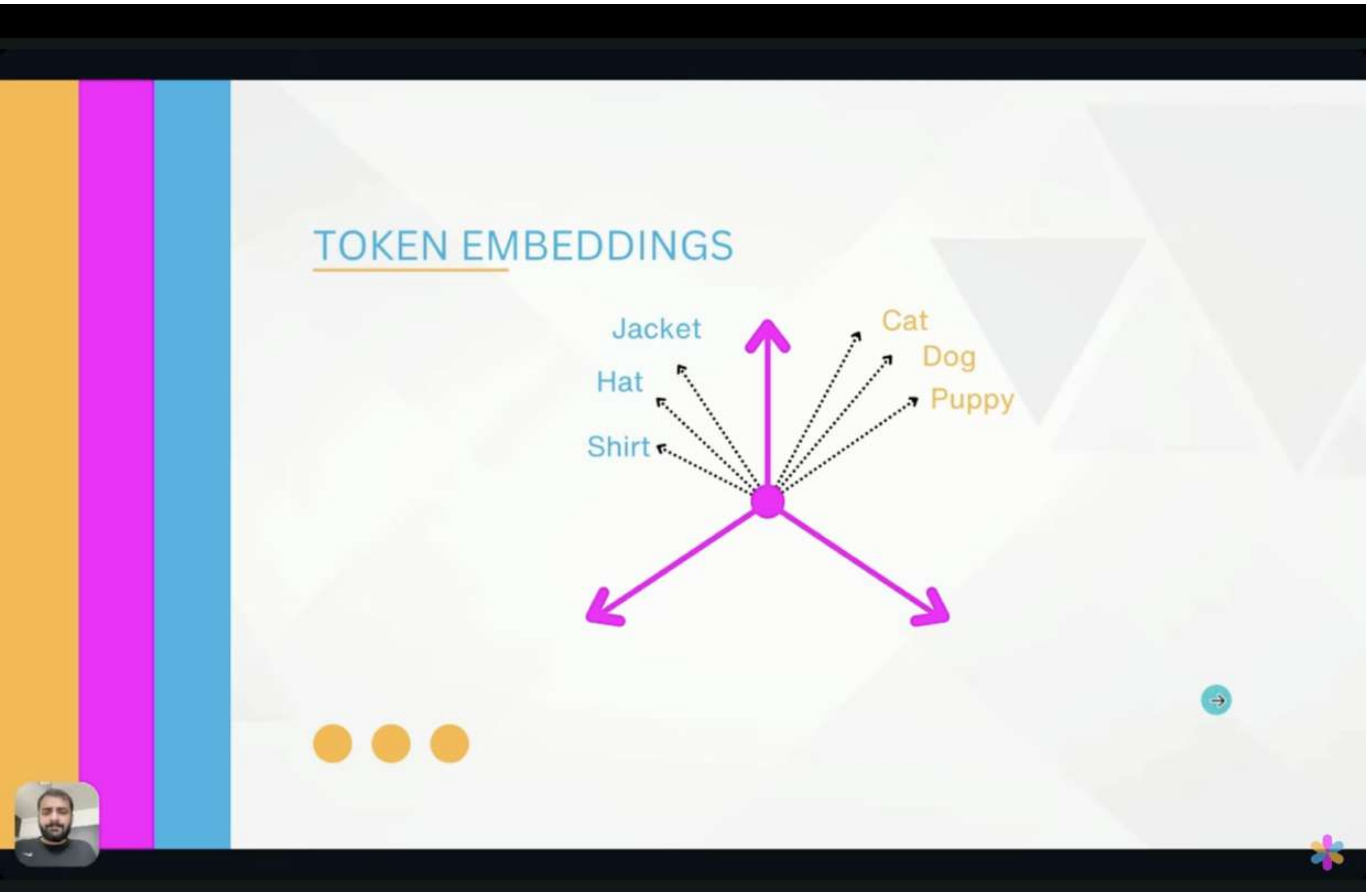


Build

LLMs from Scratch

What are token embeddings and why we need them?



TOKEN EMBEDDINGS

Jacket

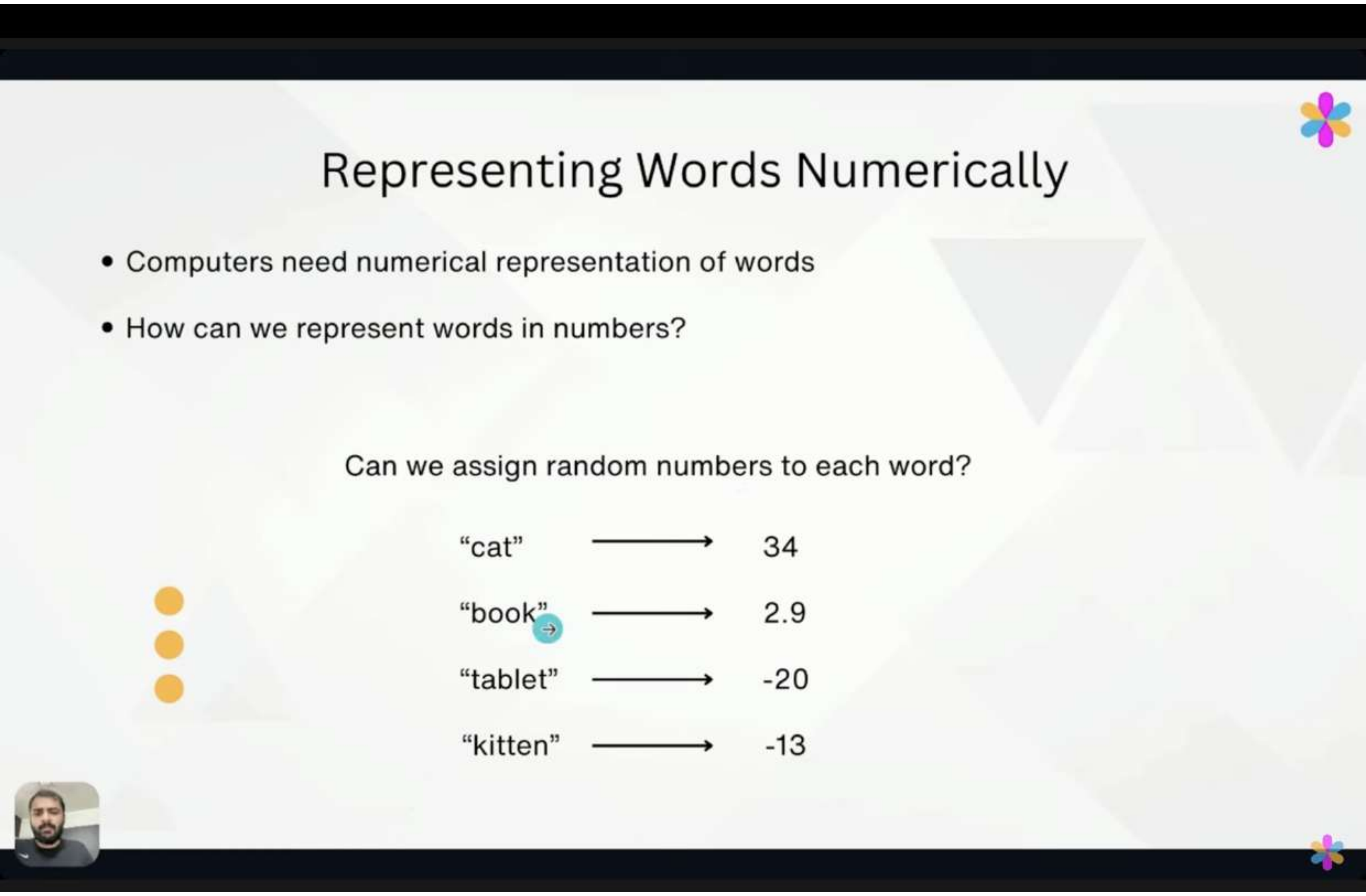
Hat

Shirt

Cat

Dog

Puppy

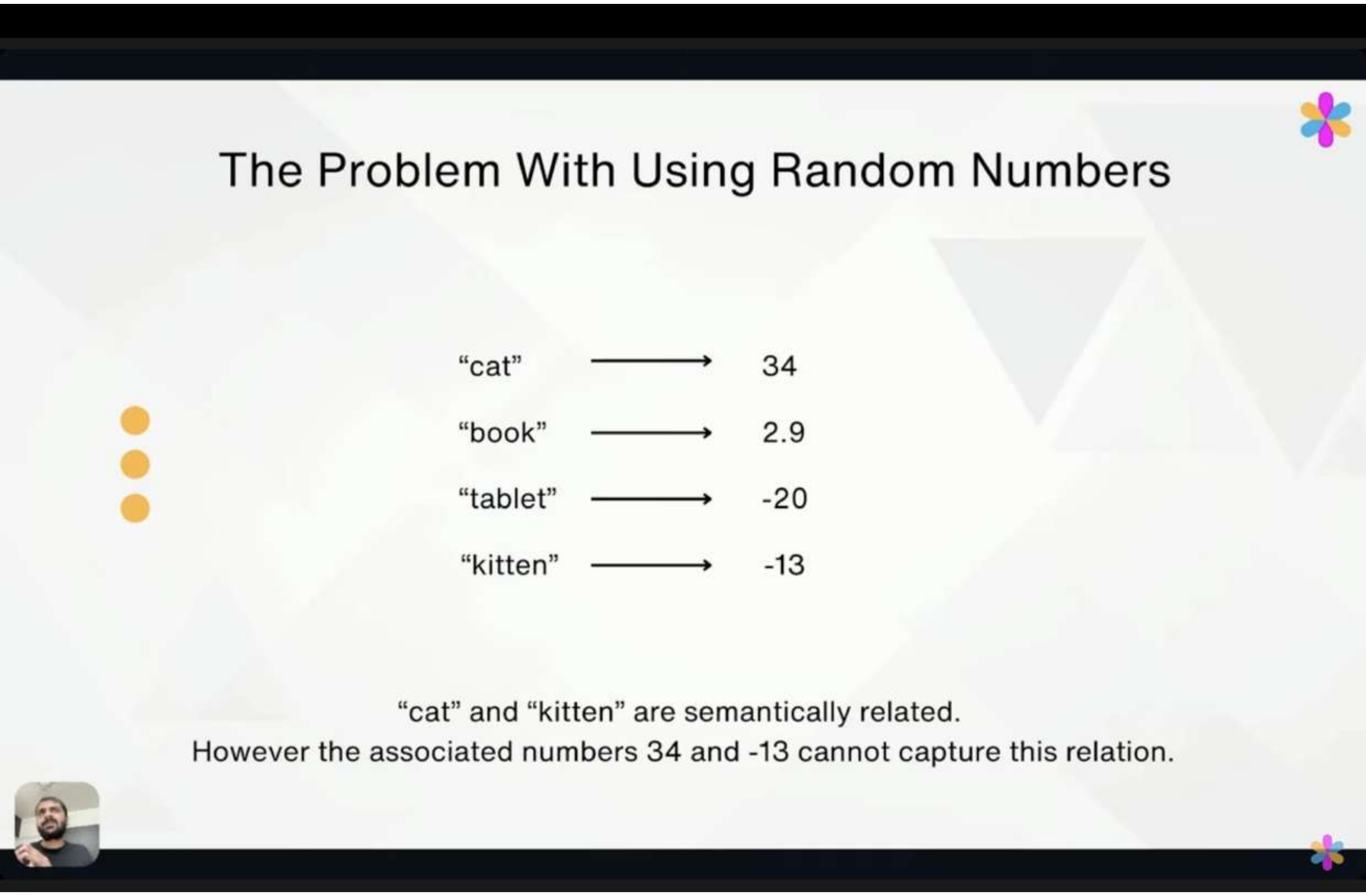


Representing Words Numerically

- Computers need numerical representation of words
- How can we represent words in numbers?

Can we assign random numbers to each word?

“cat”	→	34
“book”	→	2.9
“tablet”	→	-20
“kitten”	→	-13



The Problem With Using Random Numbers

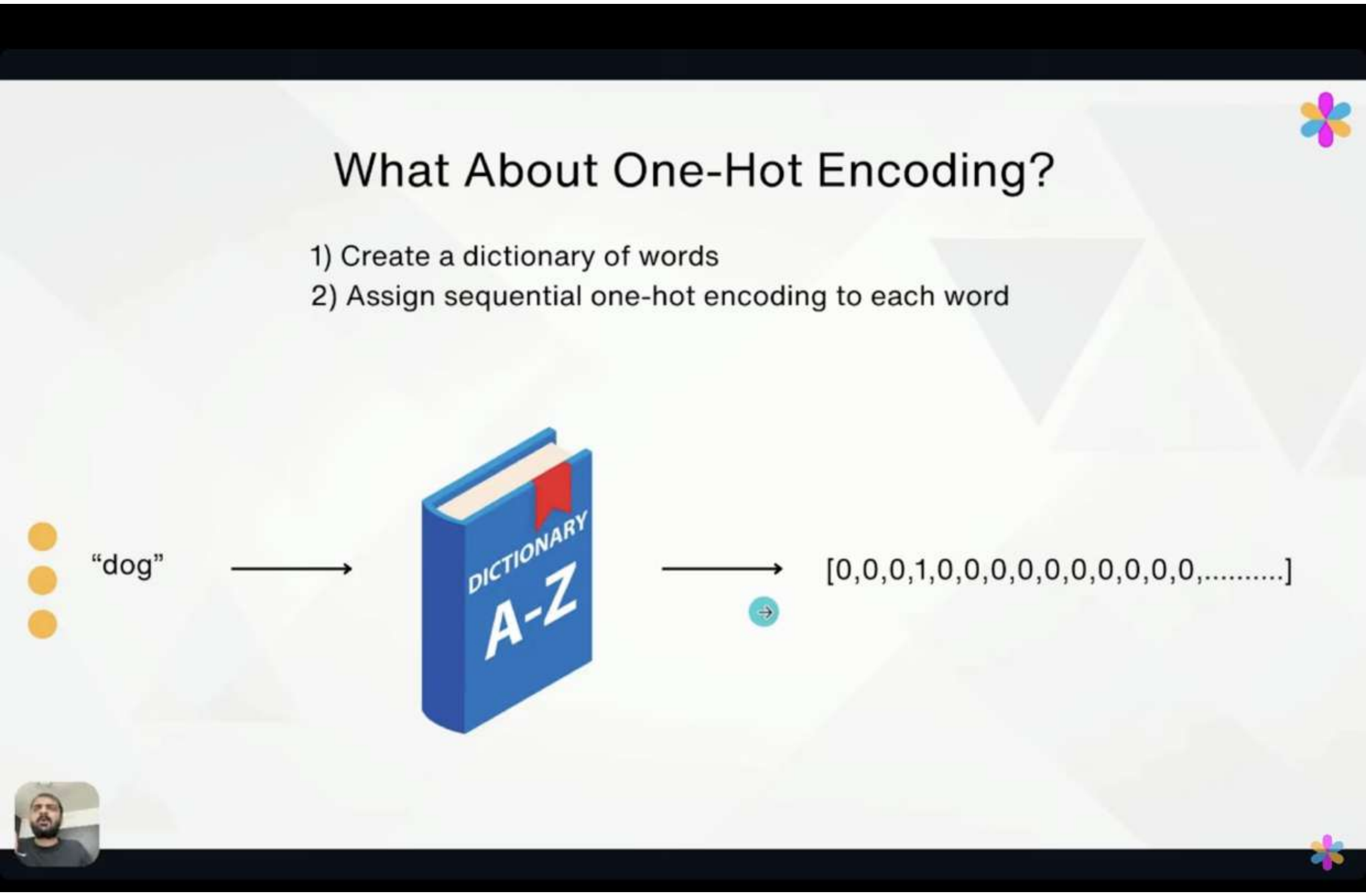
“cat” → 34

“book” → 2.9

“tablet” → -20

“kitten” → -13

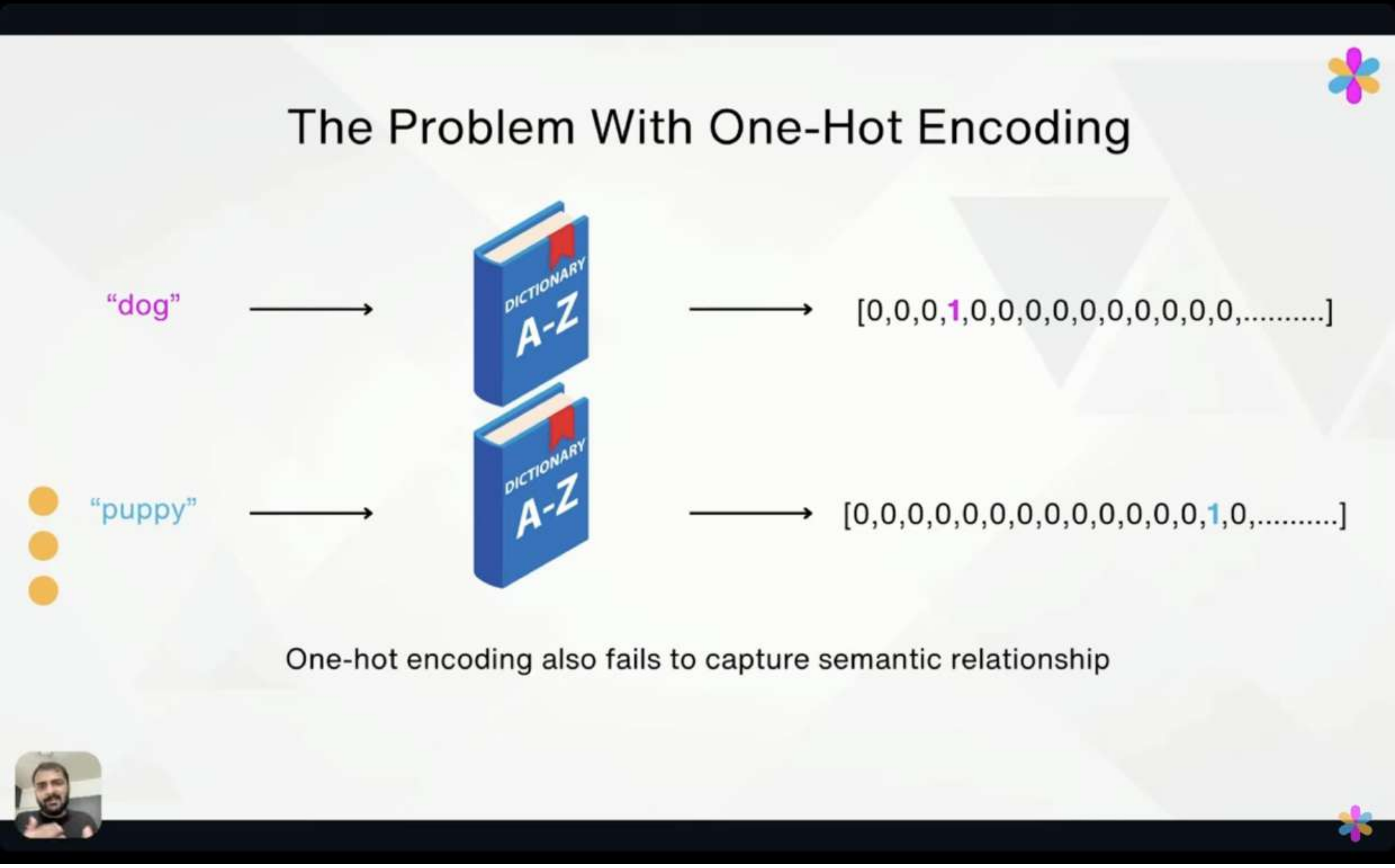
“cat” and “kitten” are semantically related.
However the associated numbers 34 and -13 cannot capture this relation.



What About One-Hot Encoding?

1) Create a dictionary of words
2) Assign sequential one-hot encoding to each word

“dog” → → [0,0,0,1,0,0,0,0,0,0,0,0,0,0,..........]

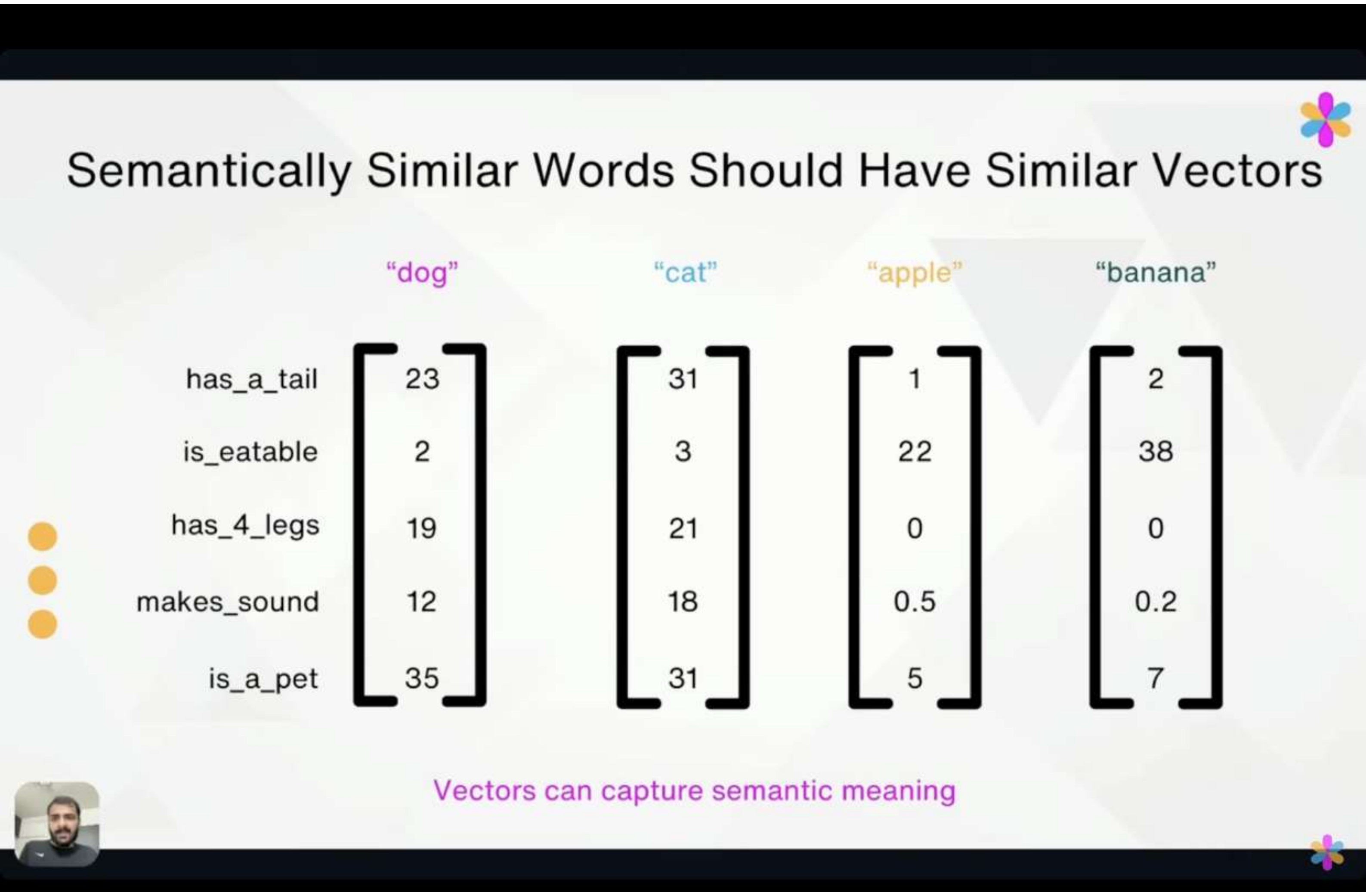


The Problem With One-Hot Encoding

“dog” → → [0,0,0,1,0,0,0,0,0,0,0,0,0,0,..........]

“puppy” → → [0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,..........]

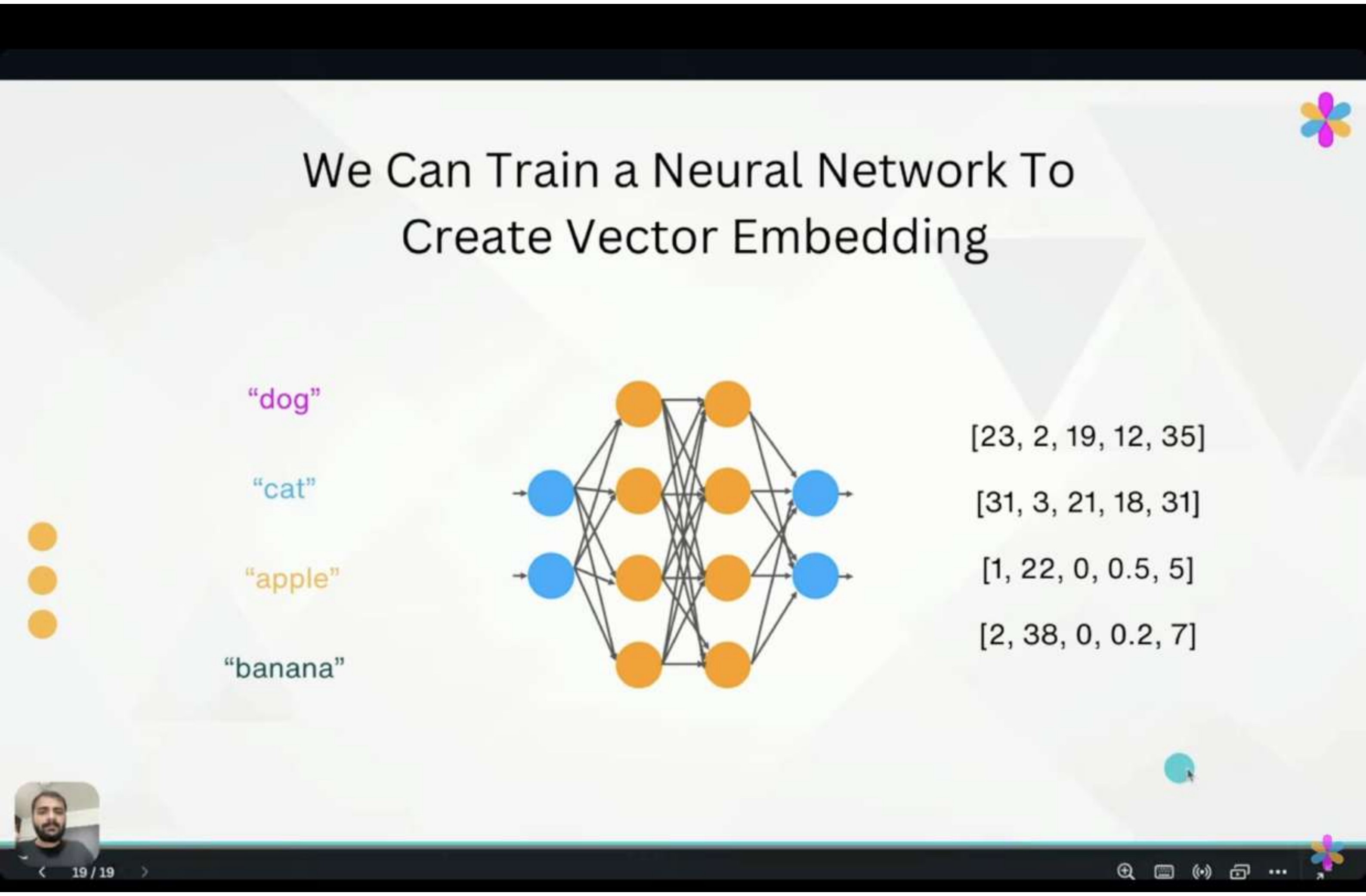
One-hot encoding also fails to capture semantic relationship



Semantically Similar Words Should Have Similar Vectors

$$\begin{array}{rcccc}
 & \text{“dog”} & \text{“cat”} & \text{“apple”} & \text{“banana”} \\
\begin{array}{r} \text{has\_a\_tail} \\ \text{is\_eatable} \\ \text{has\_4\_legs} \\ \text{makes\_sound} \\ \text{is\_a\_pet} \end{array} &
\begin{bmatrix} 23 \\ 2 \\ 19 \\ 12 \\ 35 \end{bmatrix} &
\begin{bmatrix} 31 \\ 3 \\ 21 \\ 18 \\ 31 \end{bmatrix} &
\begin{bmatrix} 1 \\ 22 \\ 0 \\ 0.5 \\ 5 \end{bmatrix} &
\begin{bmatrix} 2 \\ 38 \\ 0 \\ 0.2 \\ 7 \end{bmatrix}
\end{array}$$

Vectors can capture semantic meaning



We Can Train a Neural Network To Create Vector Embedding

"dog"

"cat"

"apple"

"banana"

[23, 2, 19, 12, 35]

[31, 3, 21, 18, 31]

[1, 22, 0, 0.5, 5]

[2, 38, 0, 0.2, 7]