

# Deep Learning : Self Attention in Transformers



राष्ट्रीय प्रौद्योगिकी संस्थान सिक्किम  
NATIONAL INSTITUTE OF TECHNOLOGY SIKKIM

**Course Instructor:**  
Dr. Bam Bahadur Sinha  
*Assistant Professor*  
*Computer Science & Engineering*  
*National Institute of Technology*  
*Sikkim*

# Contextual Word Representation



Light



Light weight



Light Blue



Light a Candle

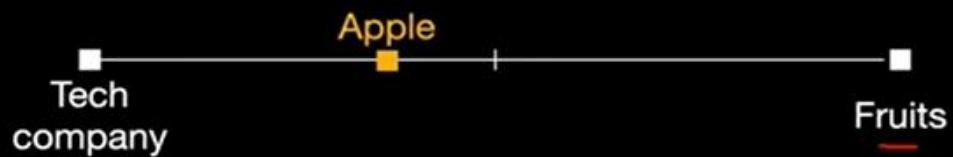
Self Attention in Transformers  
(Why we need it ?)

## Love Apple Phones

$$\text{Apple} = 0.1 \times \text{Love} + 0.5 \times \text{Apple} + 0.4 \times \text{Phones}$$

Annotations for the equation above:

- $\epsilon_{\text{Love}}$  points to  $\text{Love}$ ;  $10\% \epsilon_{\text{Love}}$  points to the coefficient  $0.1$ .
- $\epsilon_{\text{Apple}}$  points to  $\text{Apple}$ ;  $50\% \epsilon_{\text{Apple}}$  points to the coefficient  $0.5$ .
- $\epsilon_{\text{Phones}}$  points to  $\text{Phones}$ ;  $40\% \epsilon_{\text{Phones}}$  points to the coefficient  $0.4$ . This term is circled in blue.

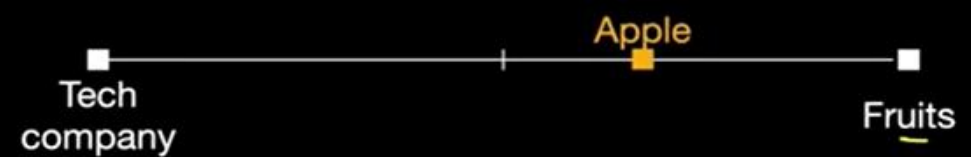


## Love Apple Juice

$$\text{Apple} = 0.1 \times \text{Love} + 0.5 \times \text{Apple} + 0.4 \times \text{Juice}$$

Annotations for the equation above:

- $\epsilon_{\text{Love}}$  points to  $\text{Love}$ .
- $\epsilon_{\text{Apple}}$  points to  $\text{Apple}$ .
- $\epsilon_{\text{Phones}}$  points to  $\text{Phones}$  in the original image, but the coefficient  $0.4$  is circled in red and labeled  $\text{Juice}$  in red.



Self Attention in Transformers  
(Why we need it?)

## Love Apple Phones

$$E'_{\text{love}} = 0.8 \times E_{\text{Love}} + 0.1 \times E_{\text{Apple}} + 0.1 \times E_{\text{Phones}}$$

$$E'_{\text{Apple}} = 0.1 \times E_{\text{Love}} + 0.5 \times E_{\text{Apple}} + 0.4 \times E_{\text{Phones}}$$

$$E'_{\text{Phones}} = 0.1 \times E_{\text{Love}} + 0.4 \times E_{\text{Apple}} + 0.5 \times E_{\text{Phones}}$$

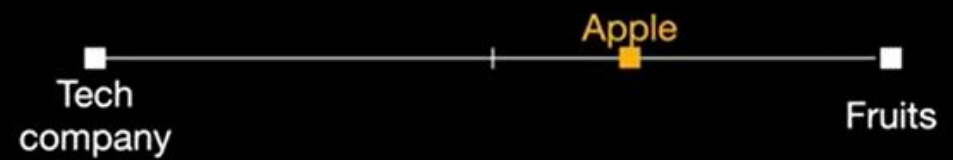


## Love Apple Juice

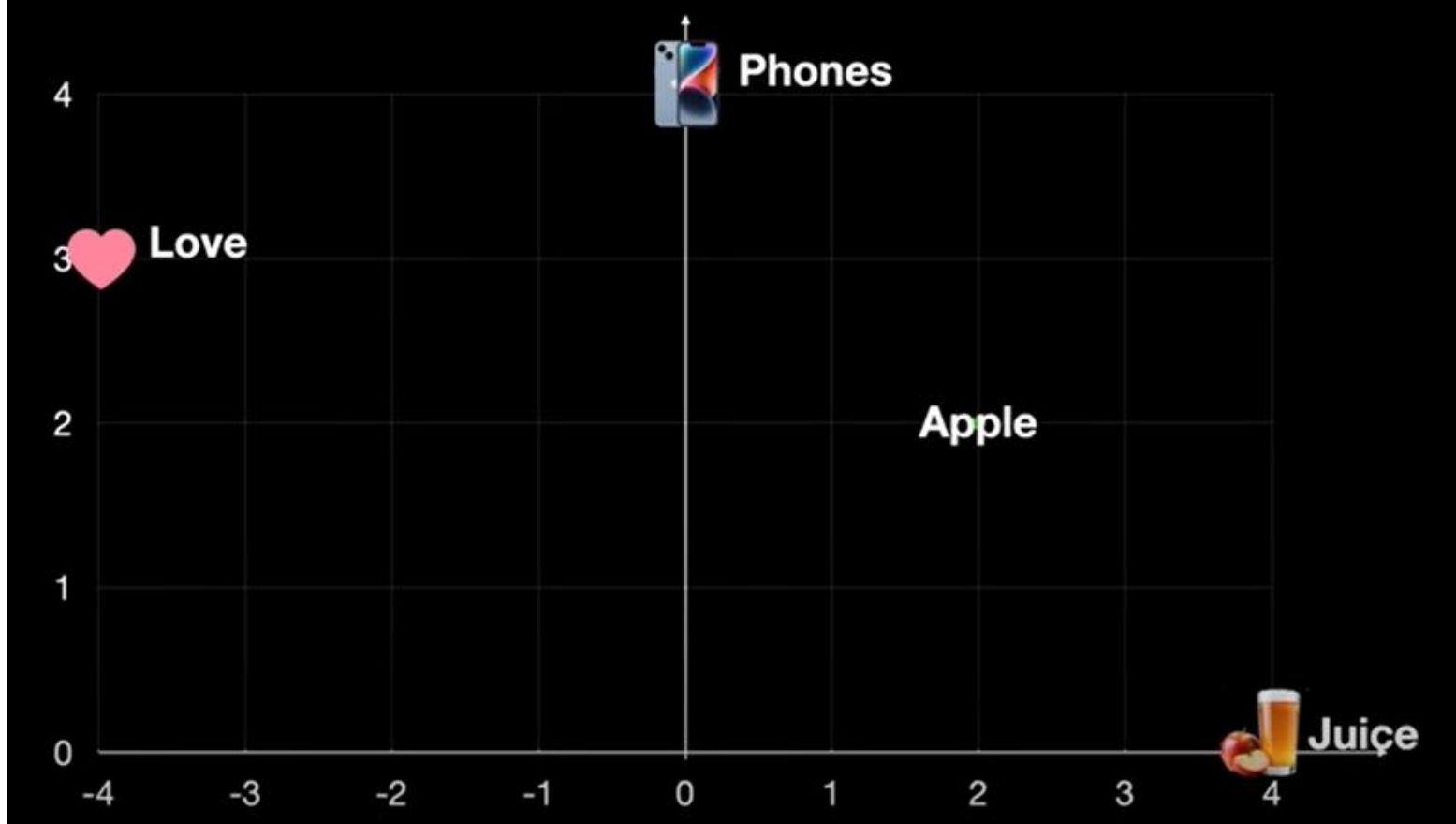
$$E'_{\text{love}} = 0.8 \times E_{\text{Love}} + 0.1 \times E_{\text{Apple}} + 0.1 \times E_{\text{Juice}}$$

$$E'_{\text{Apple}} = 0.1 \times E_{\text{Love}} + 0.5 \times E_{\text{Apple}} + 0.4 \times E_{\text{Juice}}$$

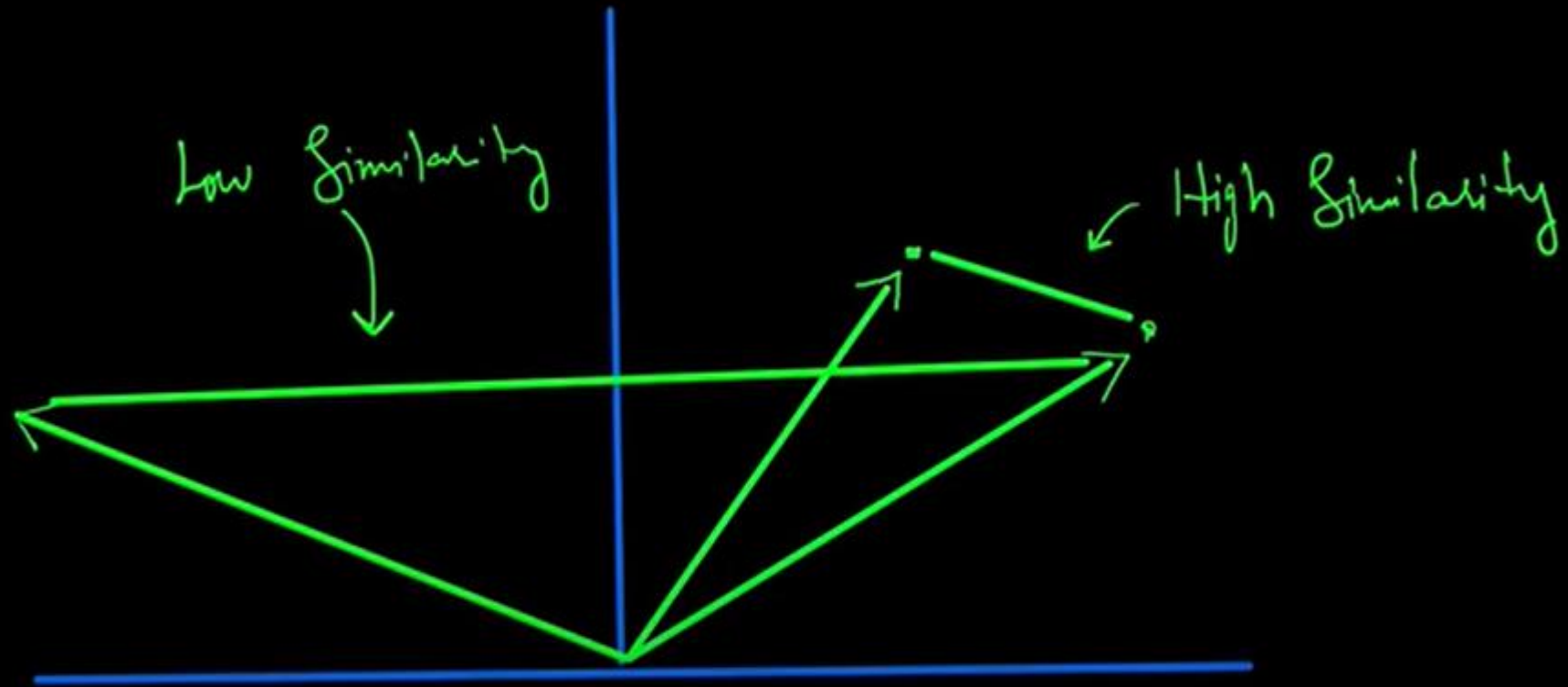
$$E'_{\text{Juice}} = 0.1 \times E_{\text{Love}} + 0.4 \times E_{\text{Apple}} + 0.5 \times E_{\text{Juice}}$$



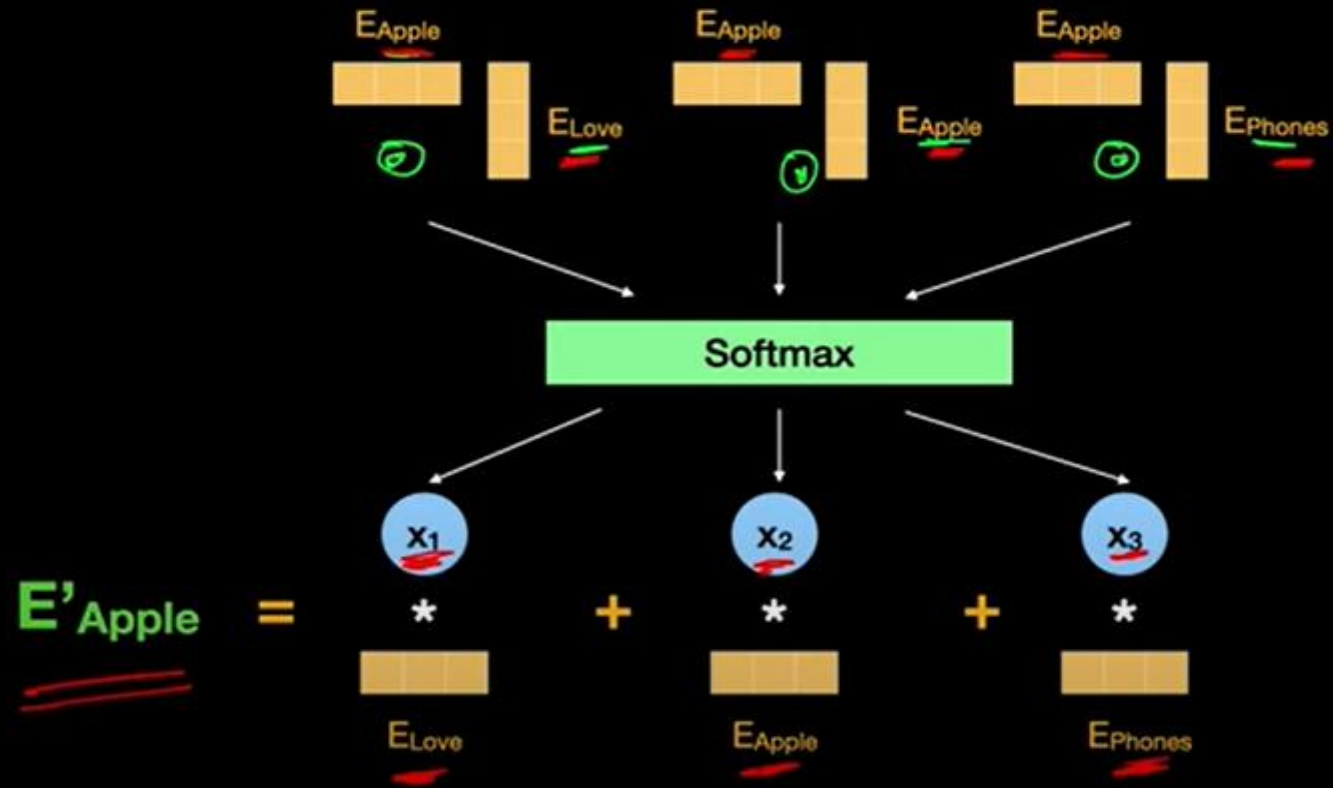
Self Attention in Transformers  
(Why we need it?)



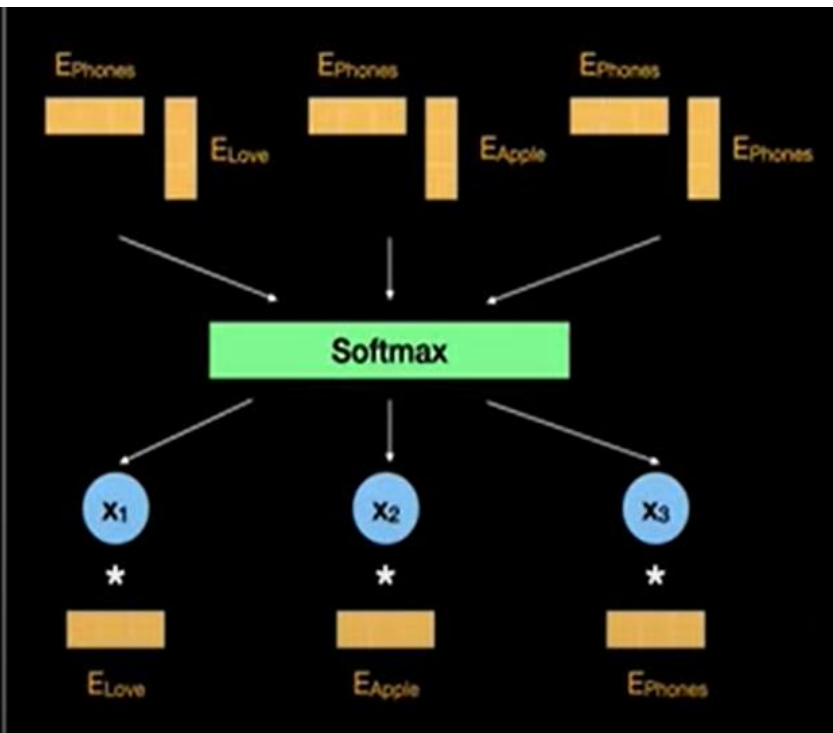
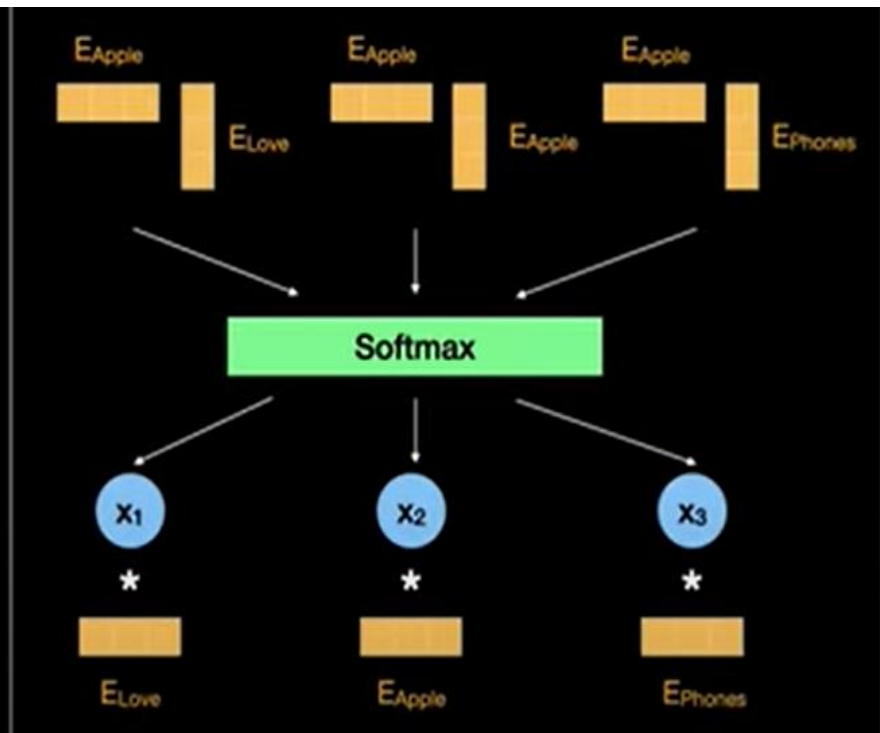
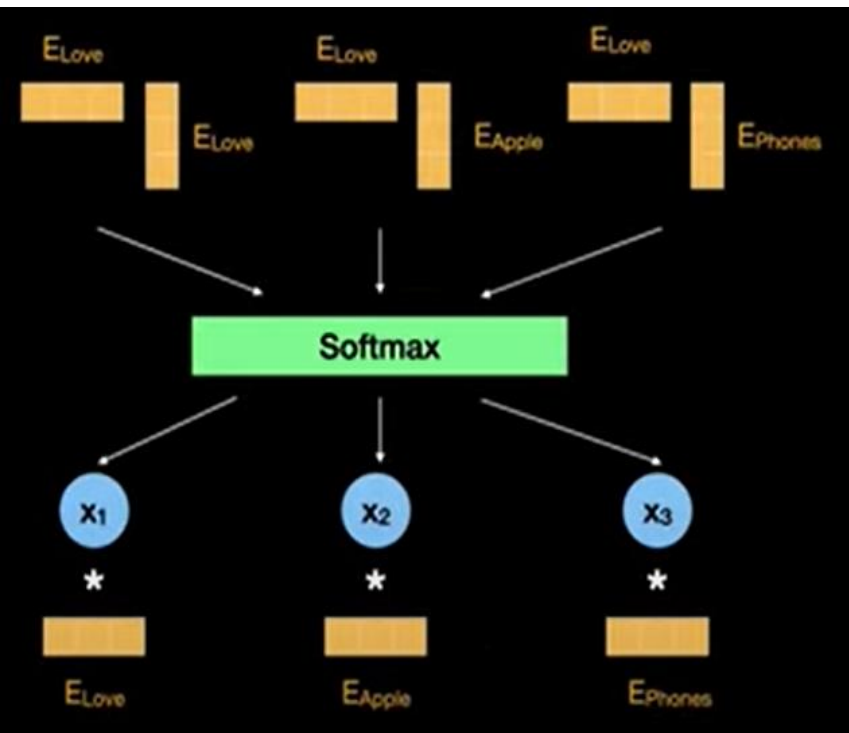
## Self Attention in Transformers (Why we need it ?)



Self Attention in Transformers  
(Dot Product Similarity)



## Self Attention in Transformers (New Word Representation)



Self Attention in Transformers  
(New Word Representation)



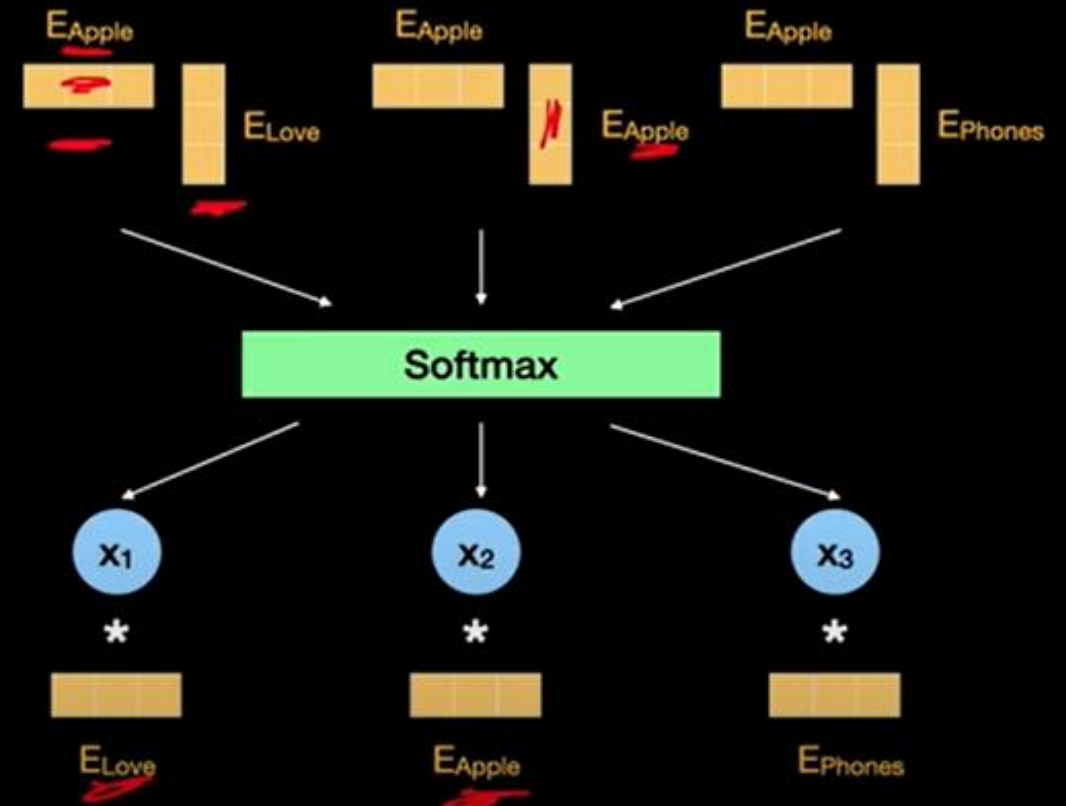
## Problem #1: No Learnable Parameters

**✗ No Task-specific training**

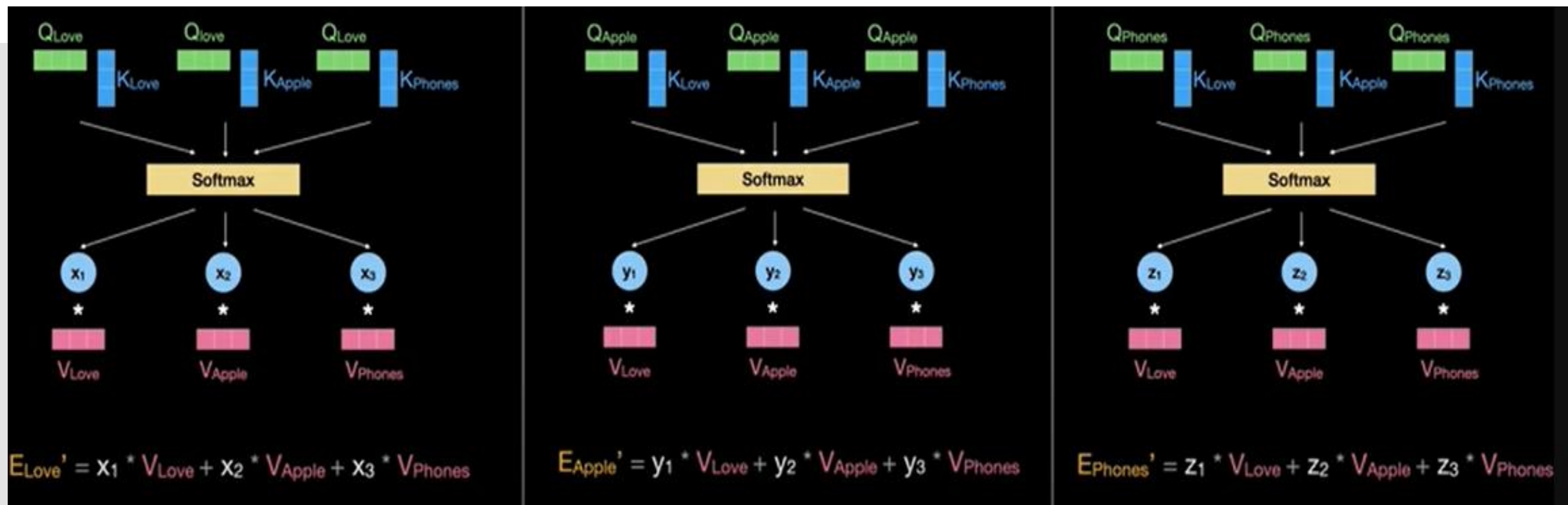


Self Attention in Transformers  
(Contextual Word Representation)

## Problem #2: Use of the same vector



Self Attention in Transformers  
(Contextual Word Representation)



## Self Attention in Transformers (Contextual Word Representation)

# Long range Dependencies

When Jake first laid eyes on the **Apple iPhone**, he wasn't expecting much. His old phone had served him well, and he wasn't one to chase the latest tech trends. But the moment he held the sleek, glassy device in his hand, there was something undeniably different about it. The polished metal edges and **smooth, responsive** screen drew him in. It felt like holding the future. He pressed the home button, and the phone lit up with a vibrant, colorful display, unlike anything he had seen before. The icons were sharp, and the touchscreen was so fluid that he couldn't resist swiping through the screens just to watch them glide by.

Before long, Jake was completely hooked. The iPhone wasn't just a phone anymore; it had become an extension of him. He loved how easily it synced with his computer, how effortlessly he could check emails, and how FaceTime made talking to family feel more personal. It wasn't just the features, though. It was the experience. Every time he picked up the phone, it was like a little piece of magic in his hands. And as the days passed, Jake realized something had shifted — he had fallen in love with this piece of technology, not just for what it could do, but for how it made his everyday life smoother, more exciting, and somehow more human.

As he started exploring, he was amazed at how intuitive everything was. Jake marvelled at the seamless transitions between apps, the pinch-to-zoom function, and the way the phone seemed to understand what he wanted to do before he did. Then he discovered the camera. He had never been one for photography, but the clarity and ease of snapping pictures with the iPhone's camera had him capturing every moment around him. The **built-in features, like visual voicemail and the music player**, made his life feel more connected and streamlined in ways he hadn't anticipated.