MACHINE LEARNING FOR SIGNAL PROCESSING

19-3-2025

Sriram Ganapathy
LEAP lab, Electrical Engineering, Indian Institute of Science,
sriramg@iisc.ac.in

Viveka Salinamakki, Varada R.
LEAP lab, Electrical Engineering, Indian Institute of Science

http://leap.ee.iisc.ac.in/sriram/teaching/MLSP25/

STORY SO FAR

Classification Problem

Decision Theory

Generative Modeling

Gaussian Modeling

Gaussian Mixture Modeling

EM algorithm

Function Modeling

Linear Models for Regression and Classification

Kernel Machines & Max-margin classifiers

Support Vector Machines

Data Representations PCA, LDA

Discriminative Modeling

Gradient Descent

Neural Networks

Learning Rules & Normalization

Deep neural architectures

STORY SO FAR

Deep neural architectures

Feed-forward Models

Convolutional Neural N/w

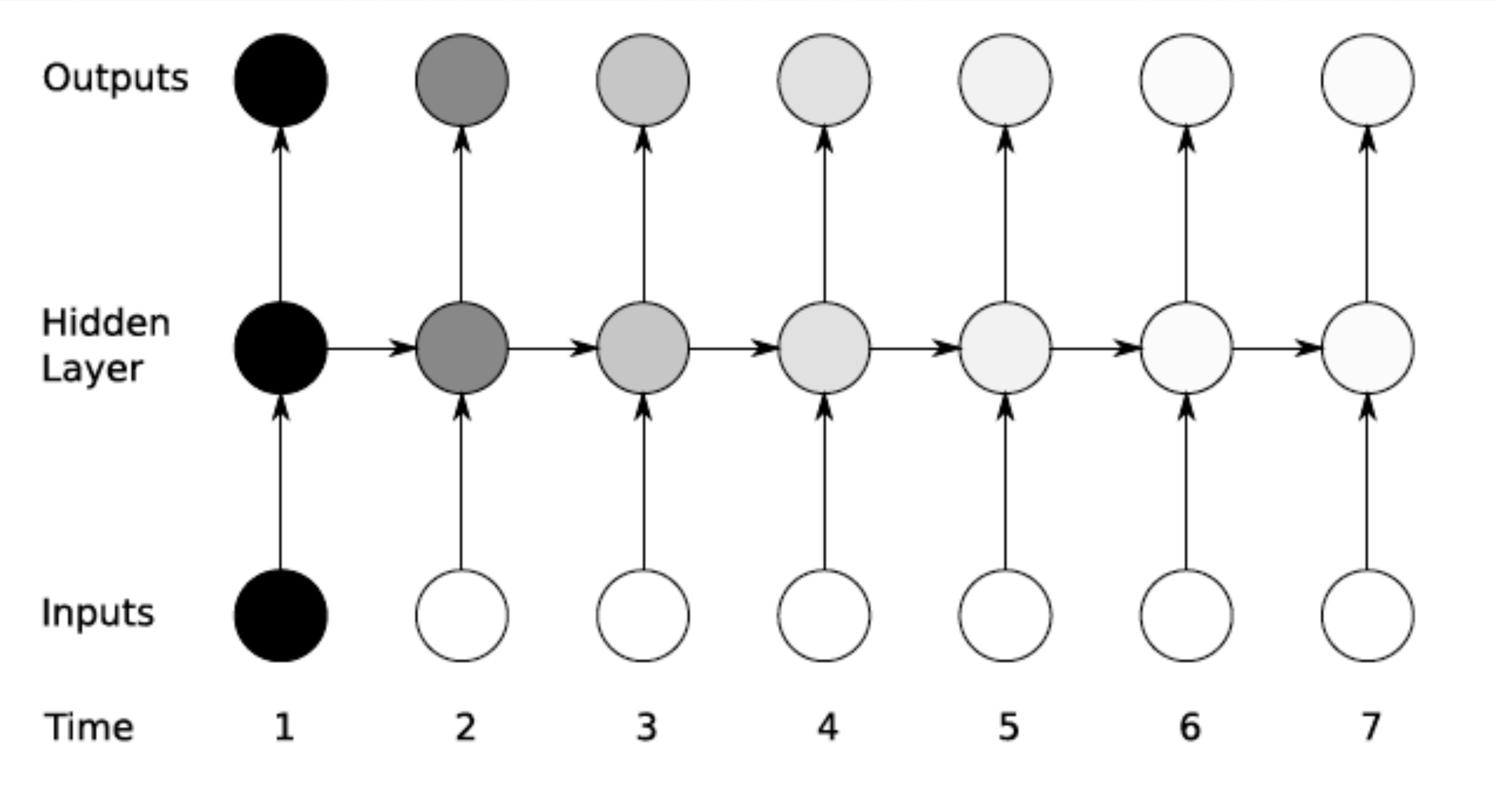
Recurrent Neural N/w

Learning with Regularization

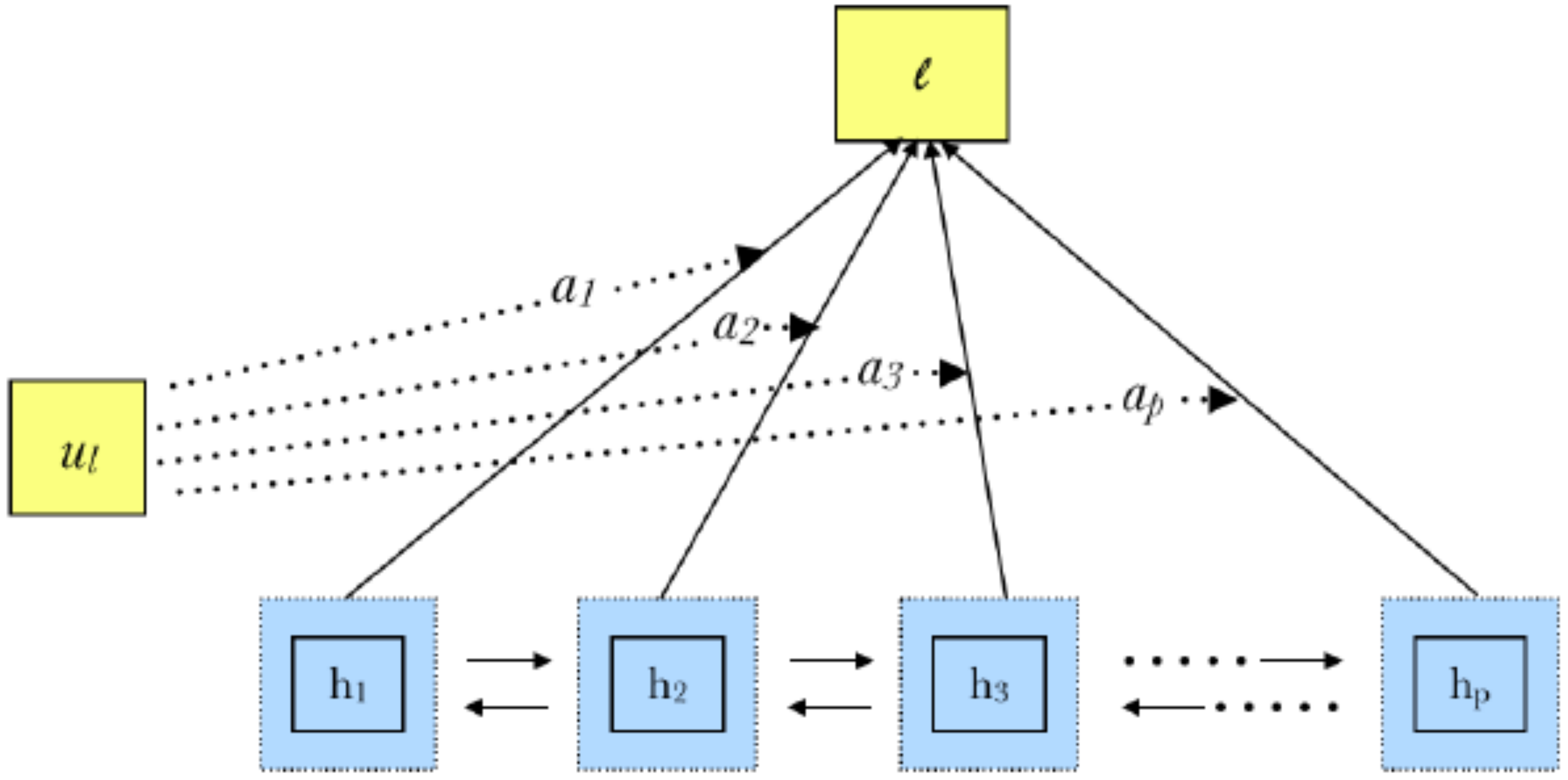
Ensembling

Attention & Transformers

Long-term Dependency Issues



Attention in LSTM Networks



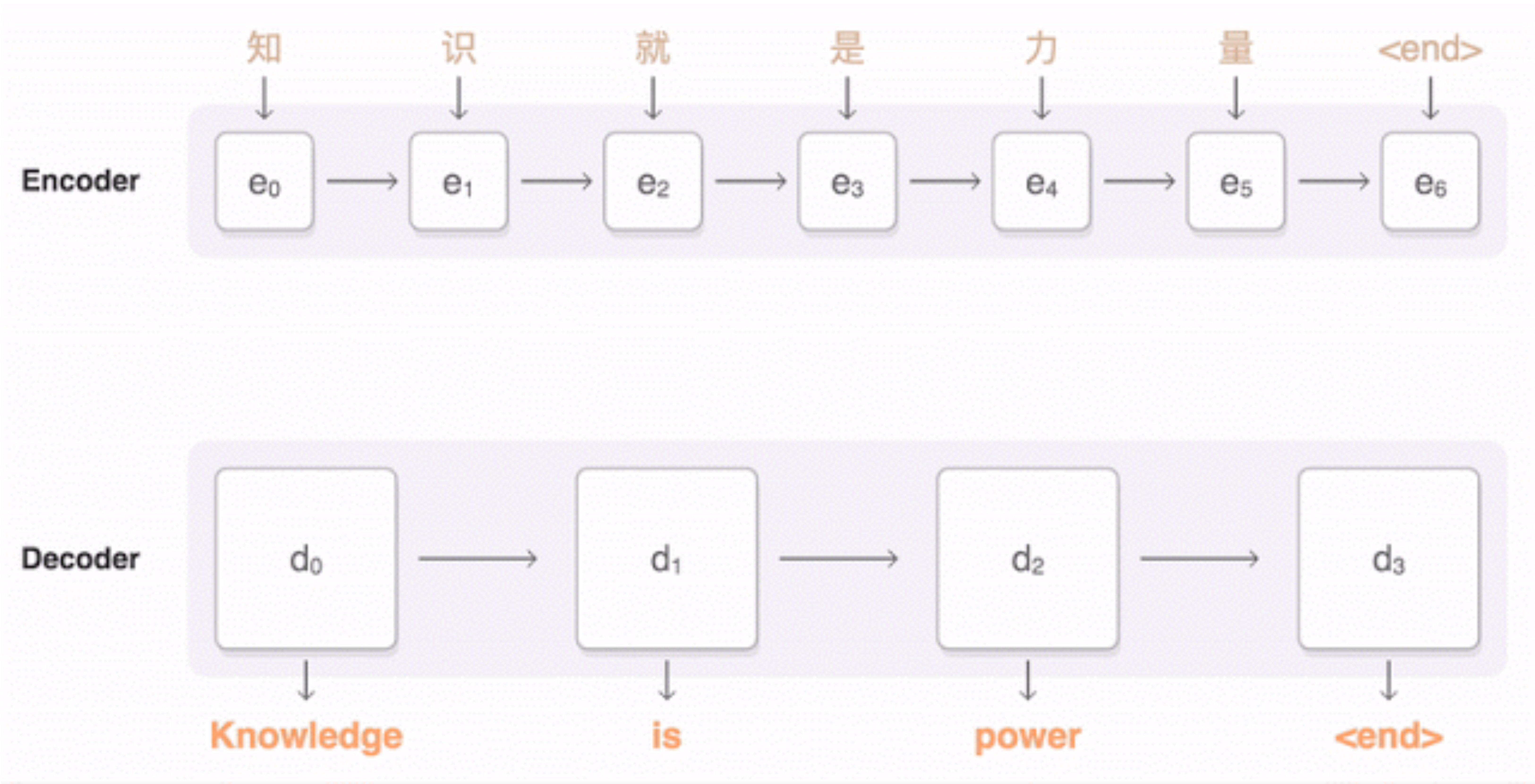
$$\mathbf{u}_t = tanh(\mathbf{W}_l \mathbf{h}_t + \mathbf{b}_l)$$

$$a_t = \frac{exp(\mathbf{u}_t^T \mathbf{u}_l)}{\sum_t exp(\mathbf{u}_t^T \mathbf{u}_l)}$$

$$\boldsymbol{l} = \sum_t a_t \mathbf{h}_t$$

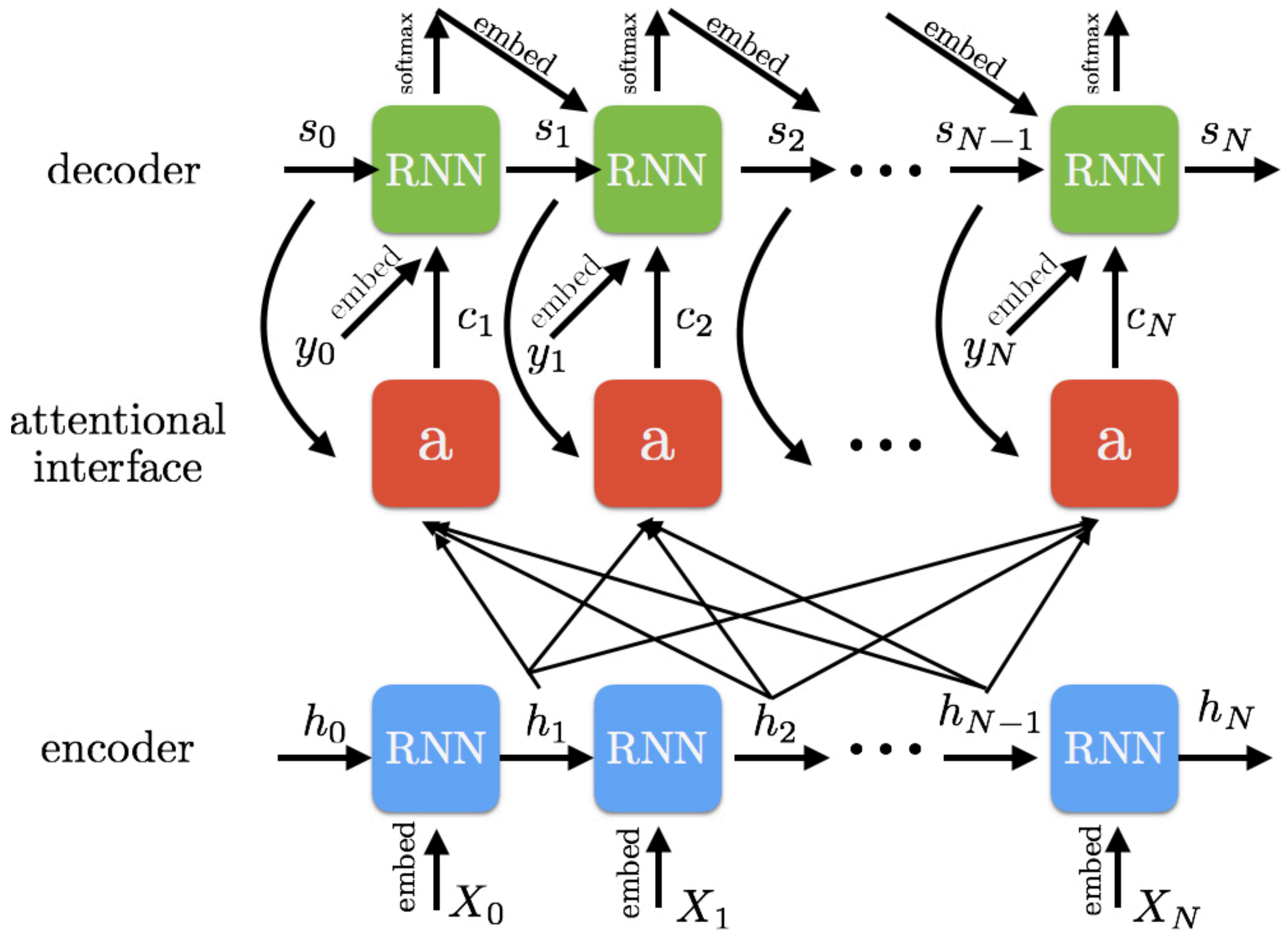
- Attentions allows a mechanism to add relevance
  - Certain regions of the audio have more importance than the rest for the task at hand.

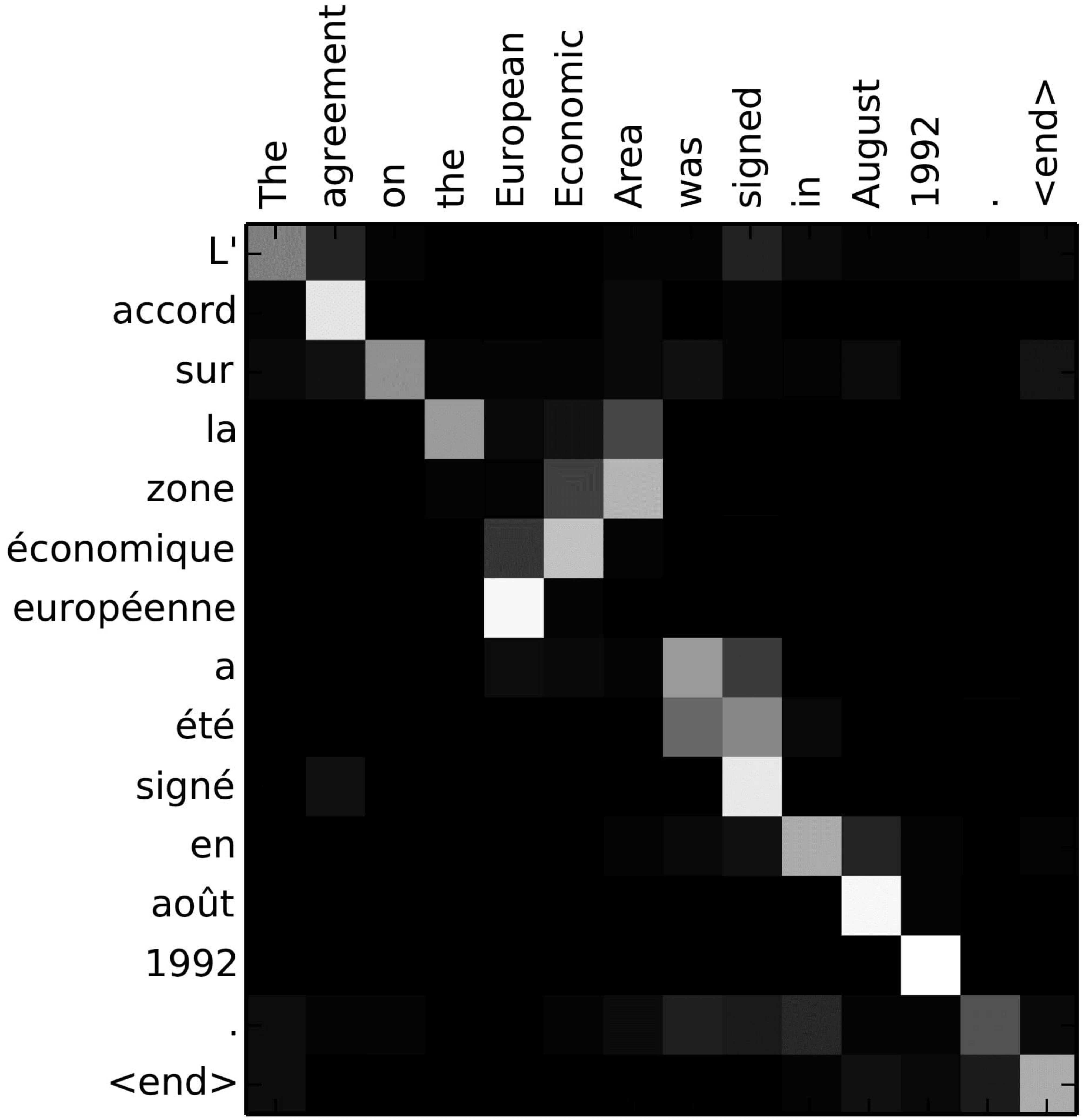
Encoder - Decoder Networks with Attention



Encoder-Decoder Attention

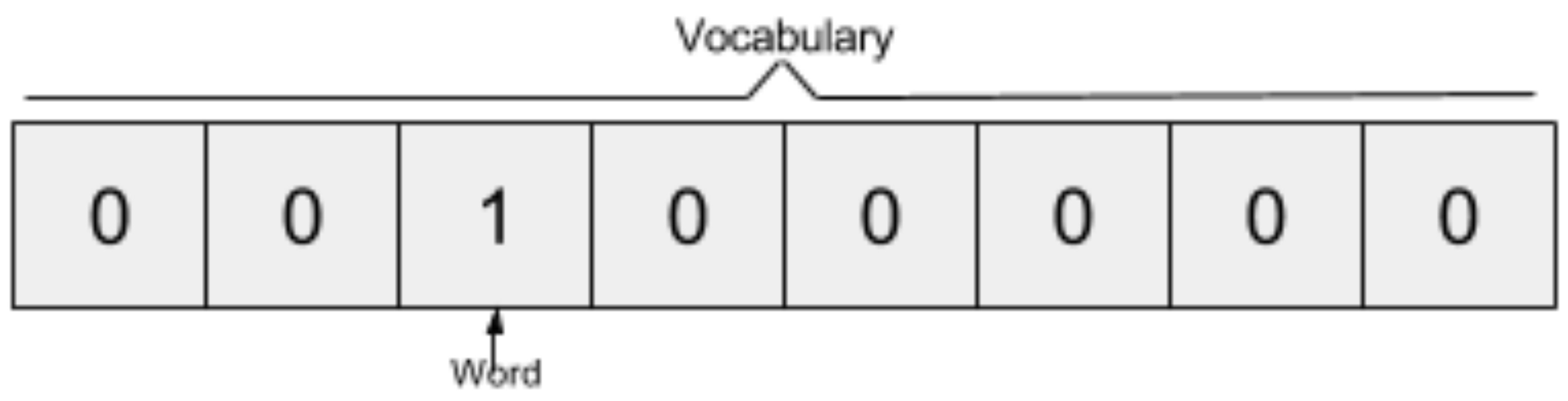
Self-Attention Models





Natural language processing - representations of text

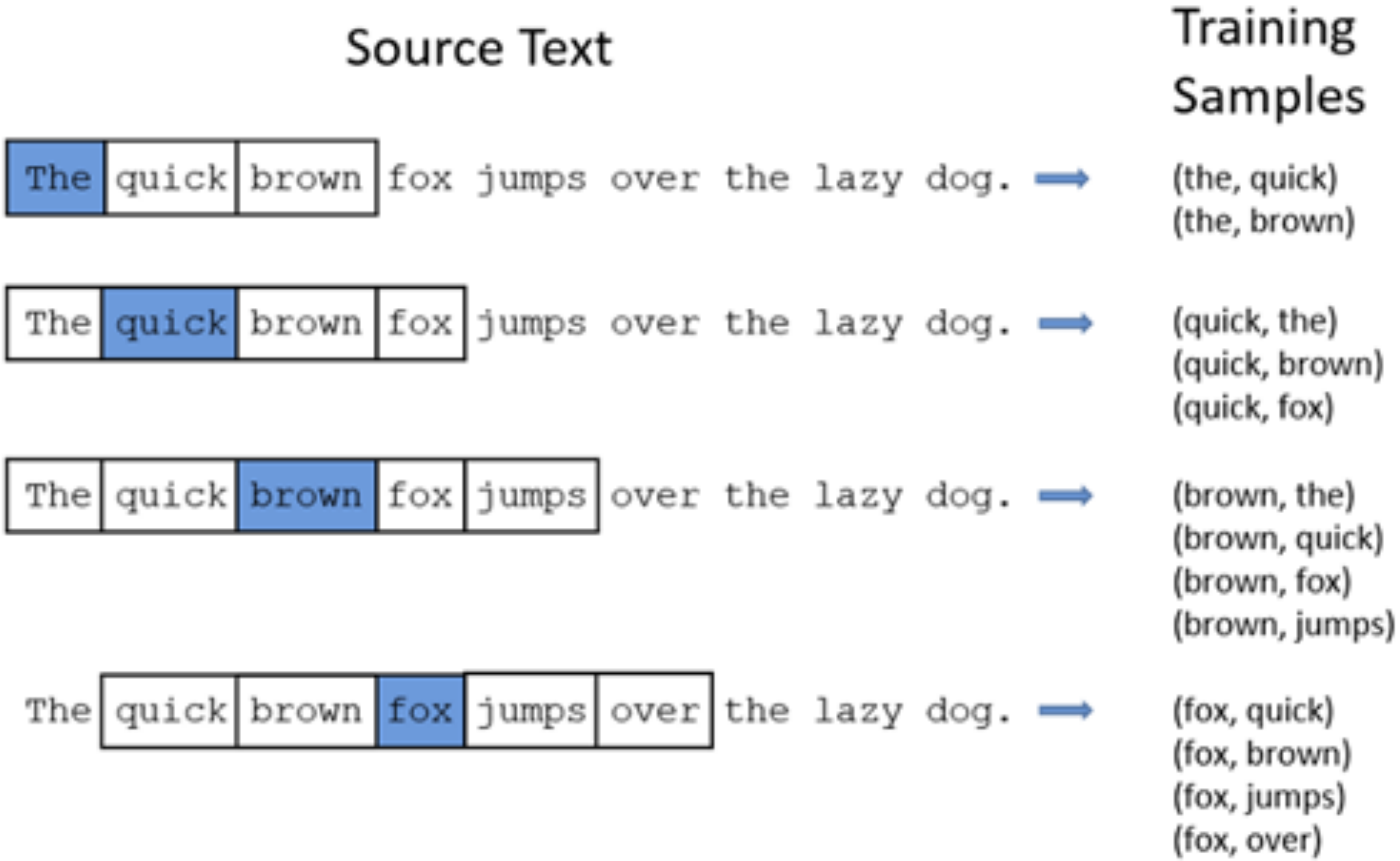
- Converting text into fixed representations in a vector space.



- Using one-hot vectors is really high dimensional.

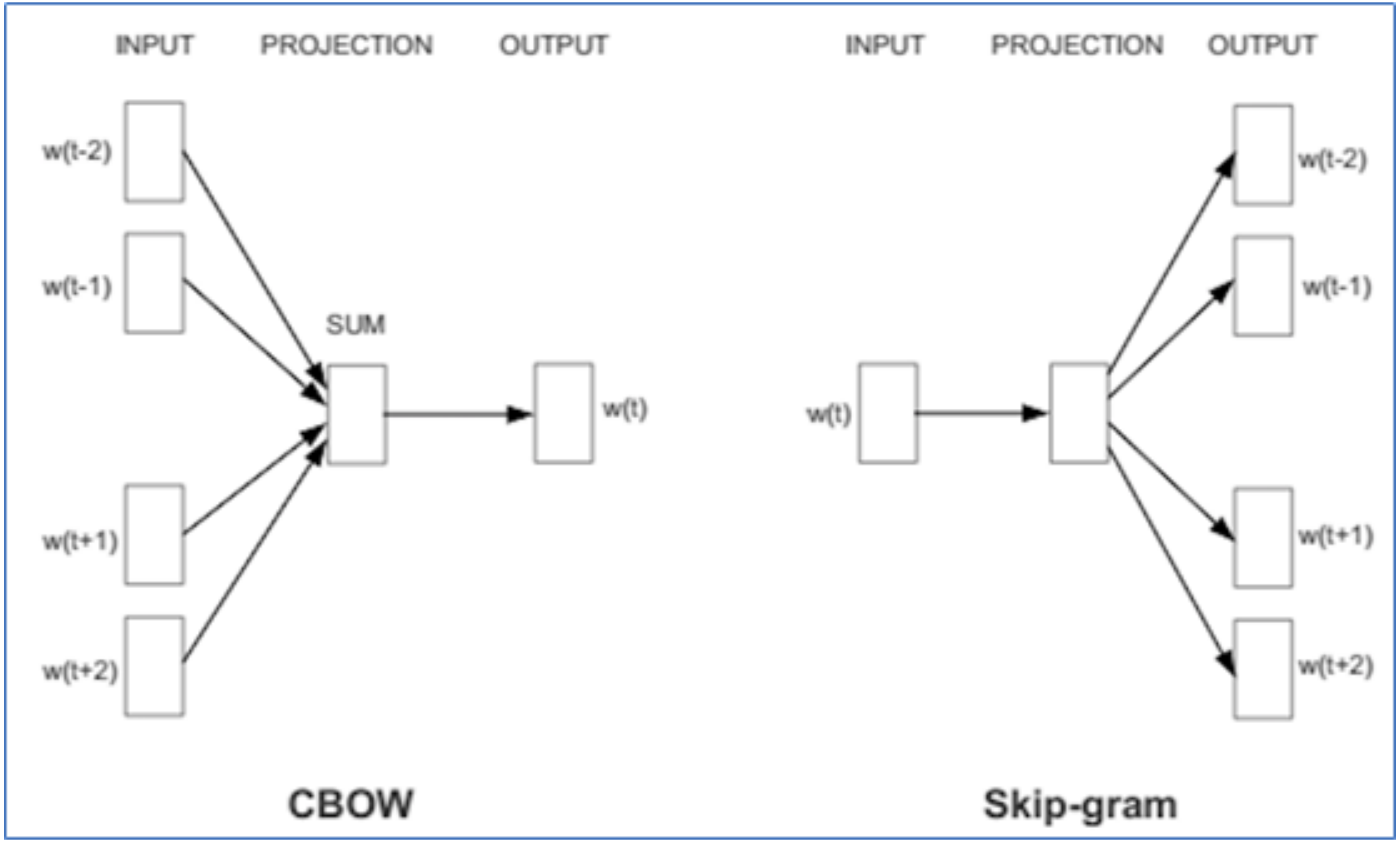
- Need a concise word representations that embeds semantics.

word2vec models as text representations

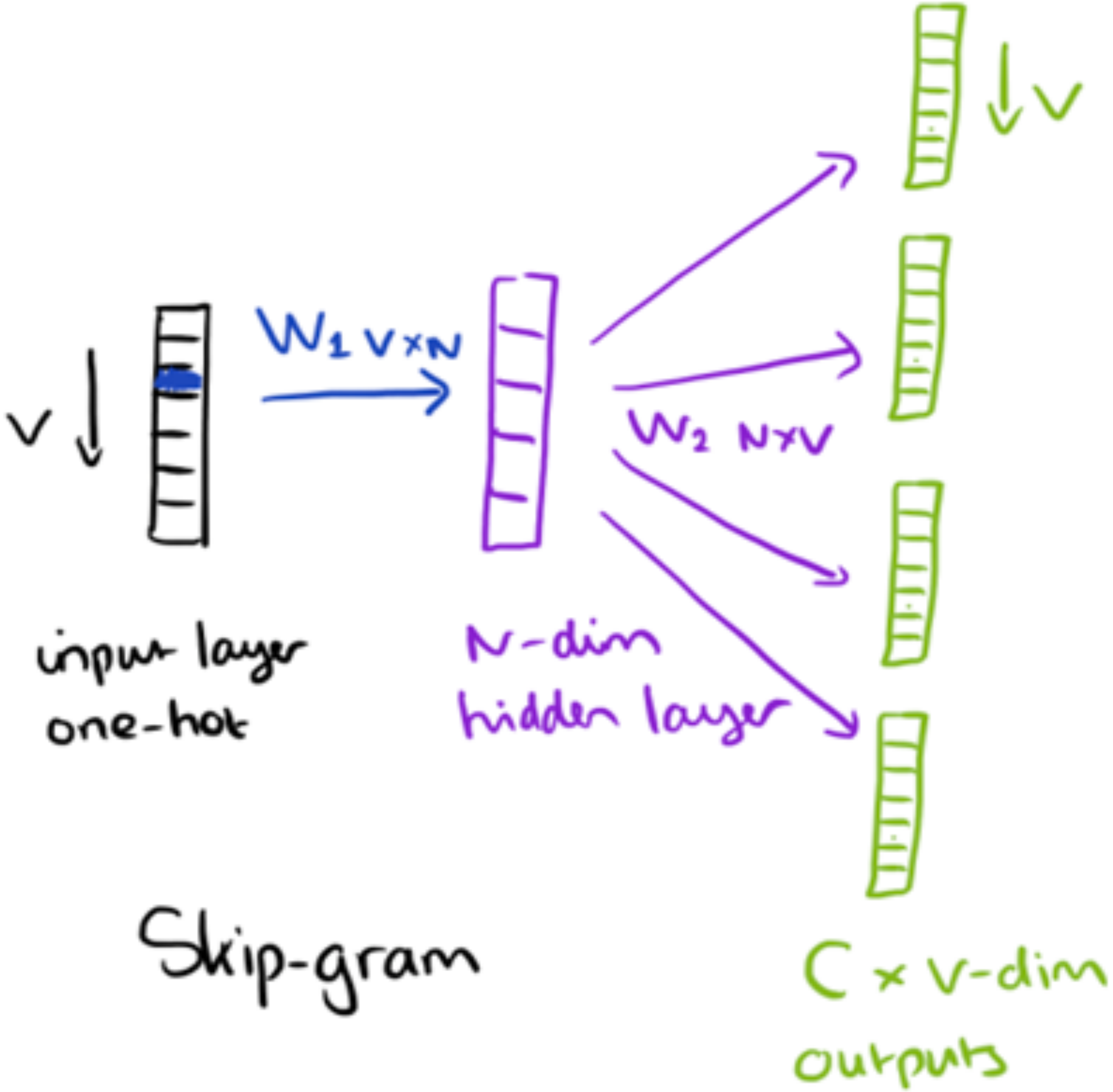


word2vec representations

word2vec models as text representations

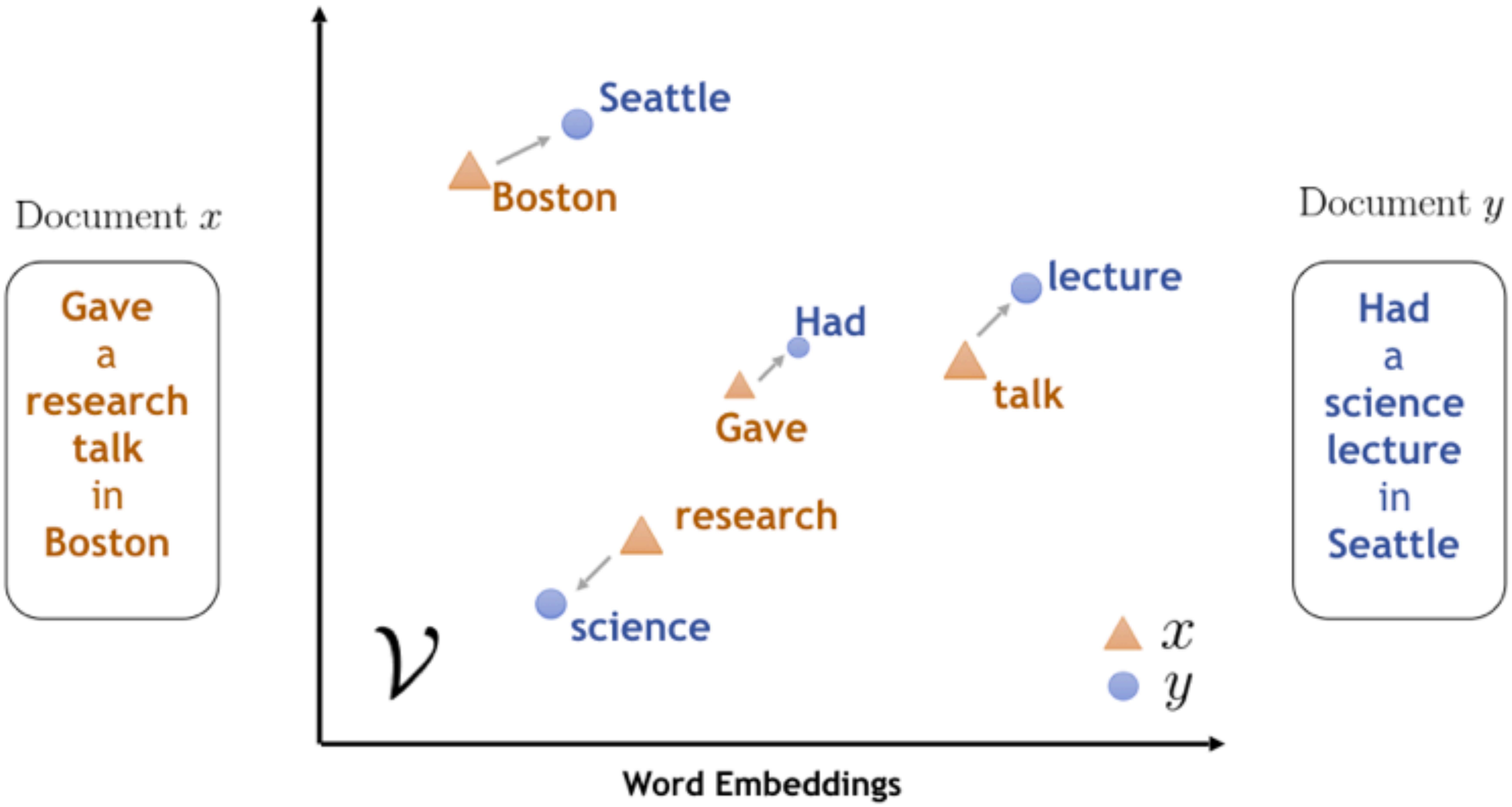


word2vec representations



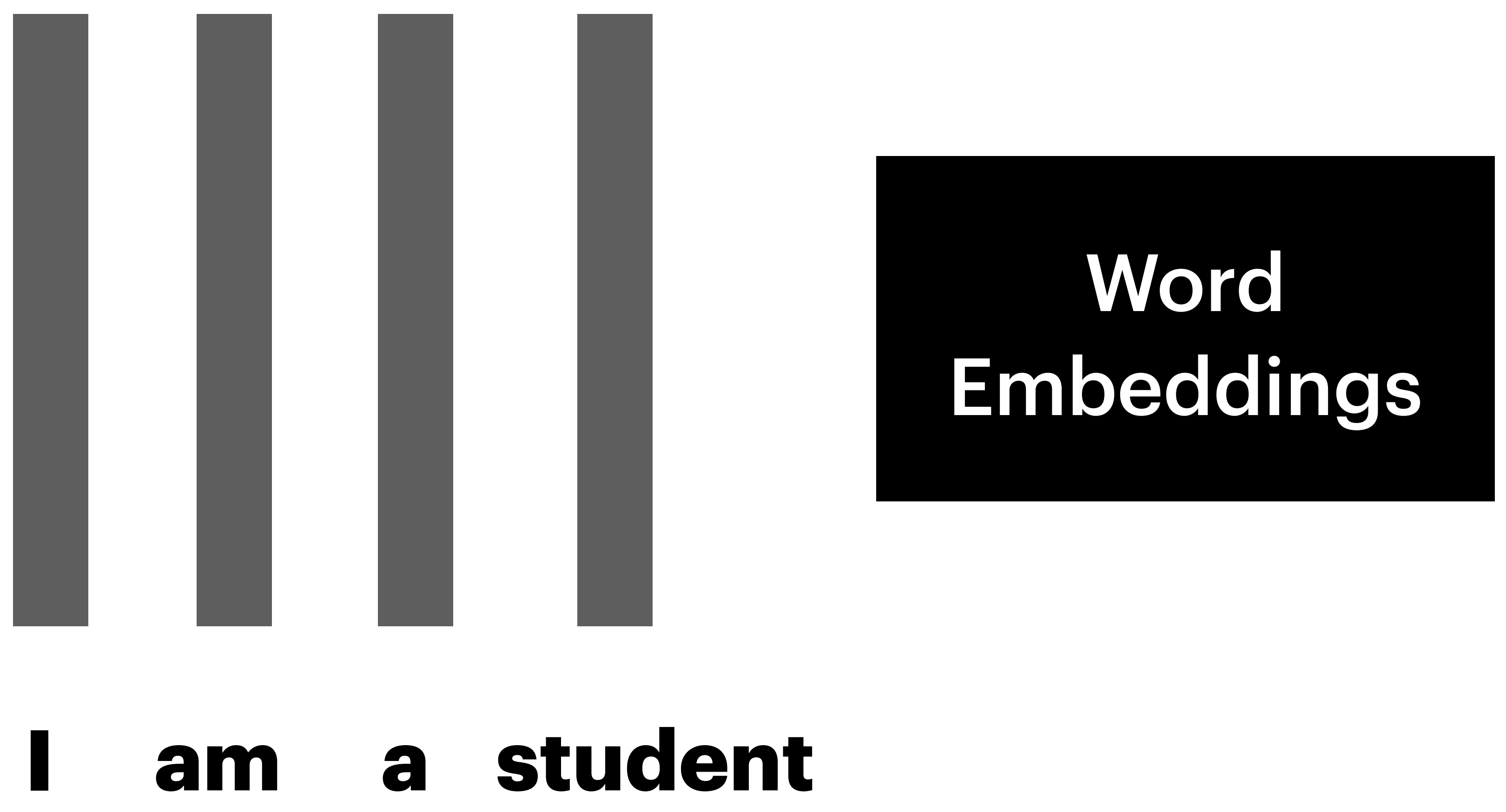
$$\begin{array}{cccc} \text{input} & W_1 & & \text{hidden}\\ 1\times V & V\times N & & 1\times N \\ \begin{bmatrix}0&1&0\end{bmatrix} & \begin{bmatrix} a&b&c&d\\ e&f&g&h\\ i&j&k&l\end{bmatrix} & = & \begin{bmatrix} e&f&g&h\end{bmatrix}\\ & W_1 & & \end{array}$$

word2vec visualization



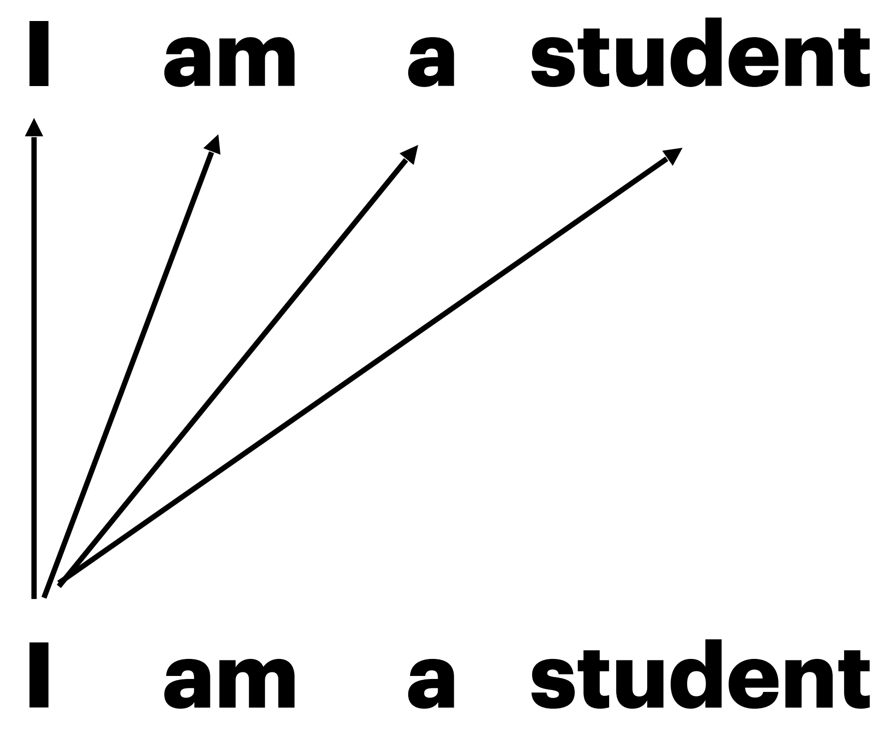
Transformers

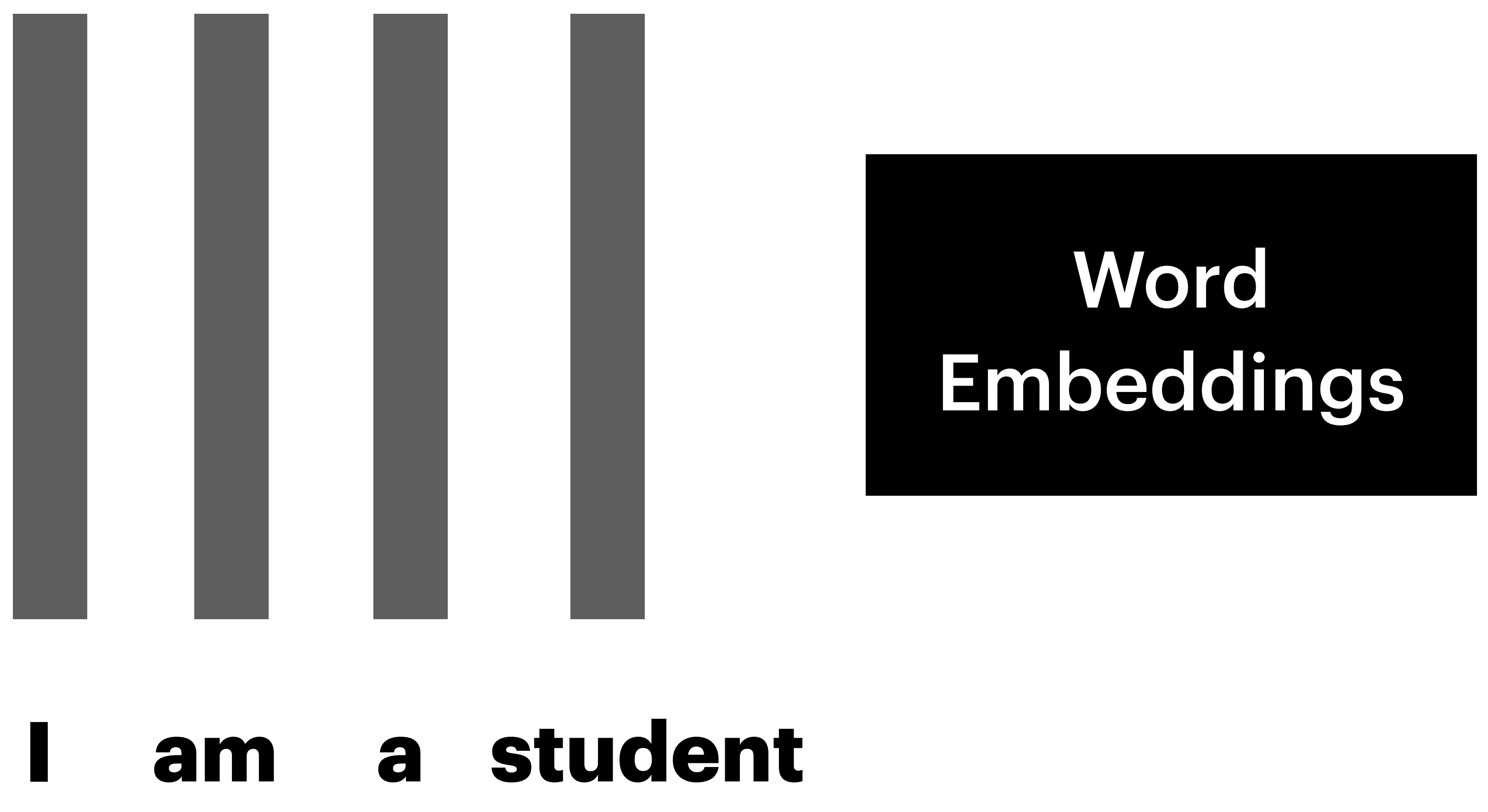
- Embedding context in sequence inputs



Transformers - self-attention

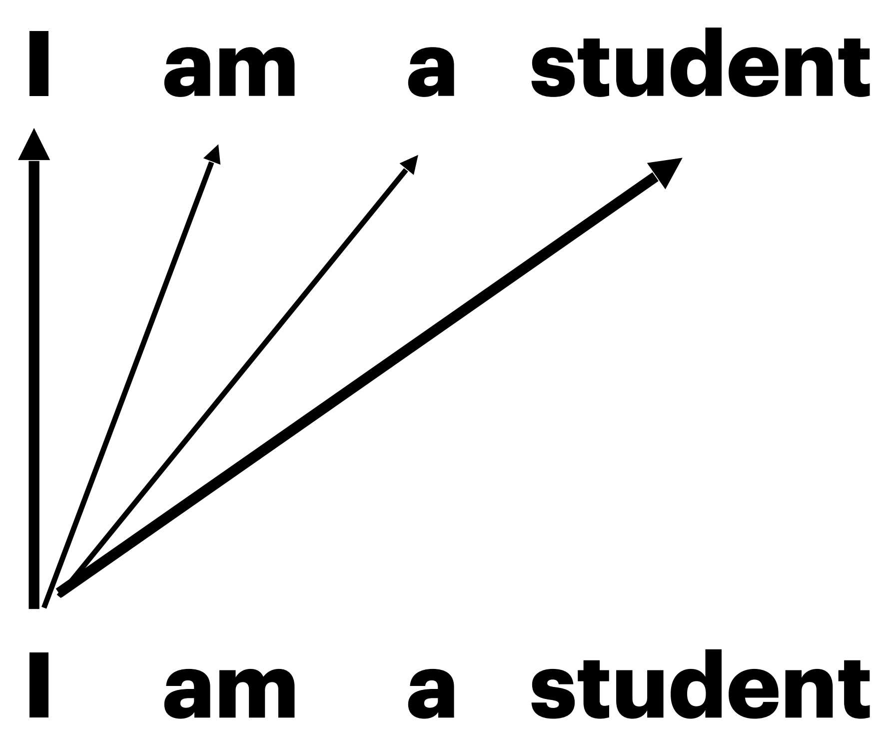
- Embedding context in sequence inputs
  - ✳ Let us take an example

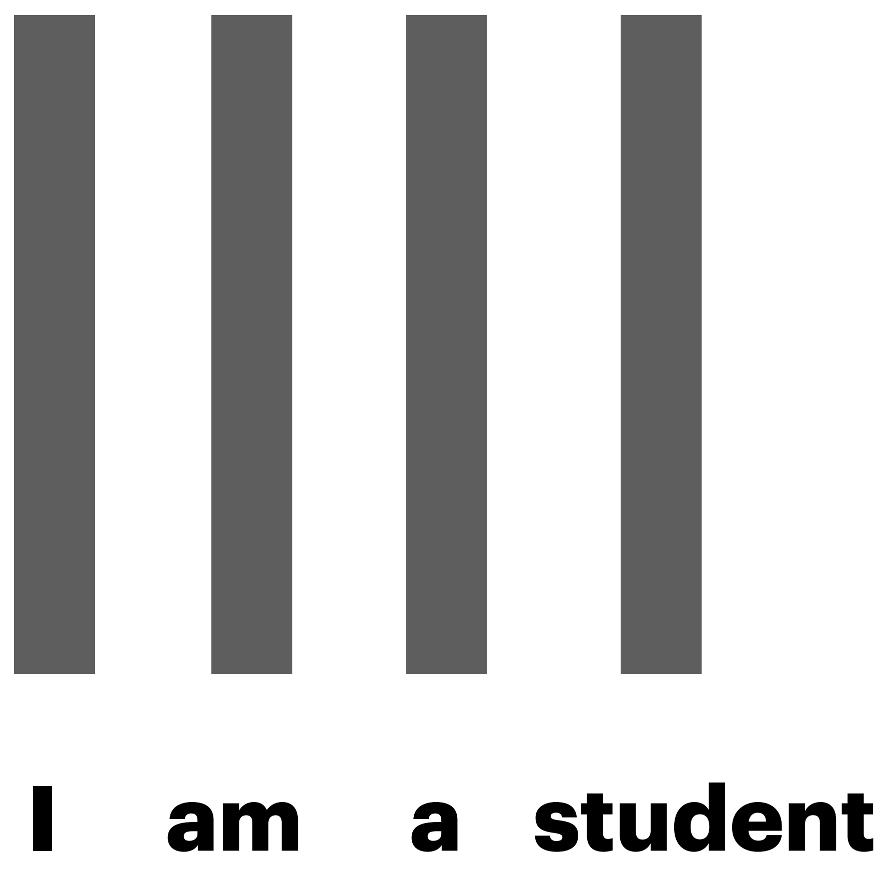




Transformers - self-attention

- Embedding context in sequence inputs
  - ✳ Let us take an example
  - ✳ Using word embeddings as the input representation





$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_T\}; \mathbf{x}_t \in \mathscr{R}^D$$

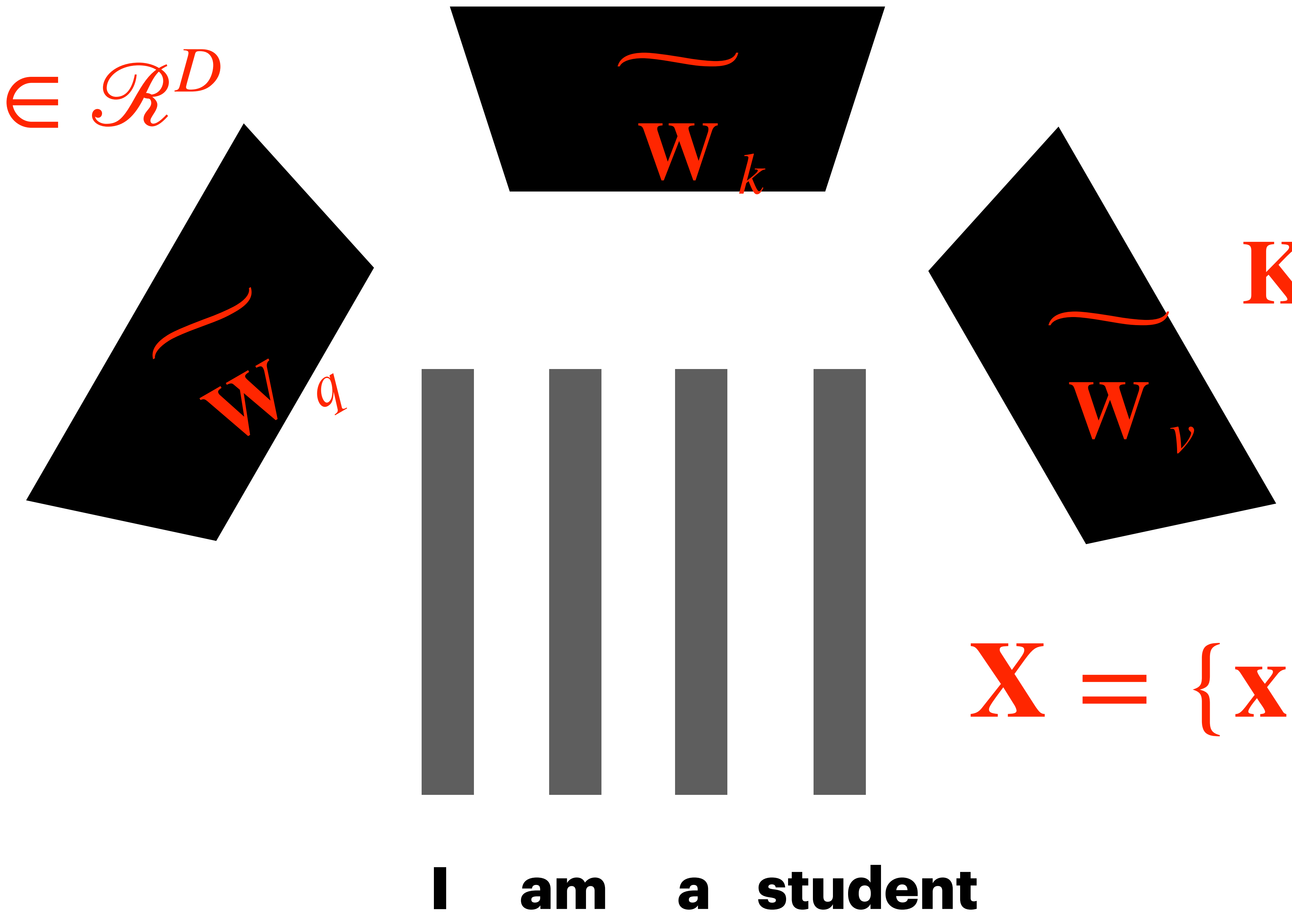
Word Embeddings

Transformers - self-attention

$$\mathbf{K} = \{\mathbf{k}_1, \mathbf{k}_2, \ldots \mathbf{k}_T\}; \mathbf{k}_t \in \mathscr{R}^D$$

$$\mathbf{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \ldots \mathbf{x}_T\}; \mathbf{q}_t \in \mathscr{R}^D$$

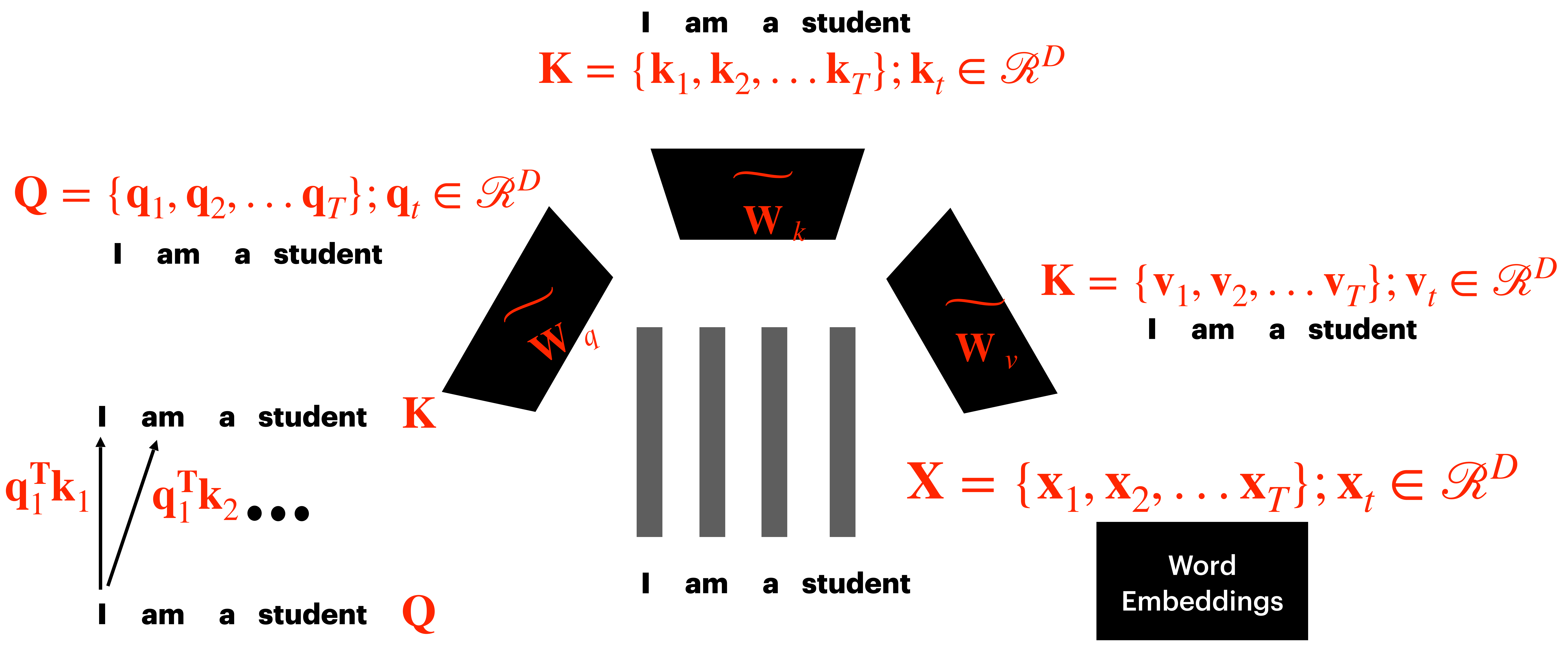
$$\mathbf{K} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_T\}; \mathbf{v}_t \in \mathscr{R}^D$$



$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_T\}; \mathbf{x}_t \in \mathscr{R}^D$$

Word Embeddings

Transformers - self-attention



Transformers - self-attention

I am a student

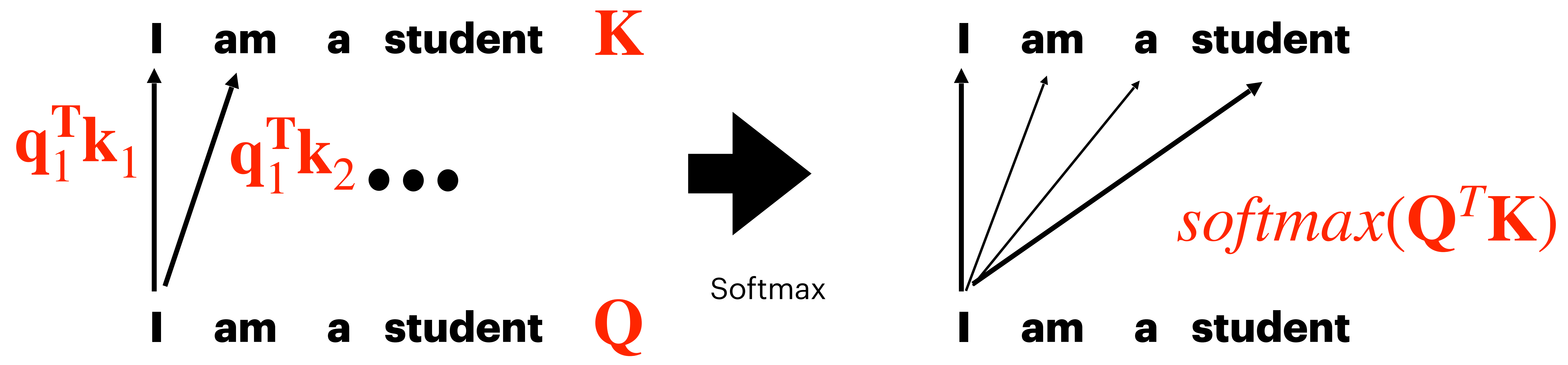
$$\mathbf{K} = \{\mathbf{k}_1, \mathbf{k}_2, \ldots \mathbf{k}_T\}; \mathbf{k}_t \in \mathscr{R}^D$$

I am a student

$$\mathbf{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \ldots \mathbf{q}_T\}; \mathbf{q}_t \in \mathscr{R}^D$$

$$\mathbf{K} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_T\}; \mathbf{v}_t \in \mathscr{R}^D$$

I am a student



Pics taken from : https://jalammar.github.io/illustrated-transformer/

Transformers - self-attention

I am a student

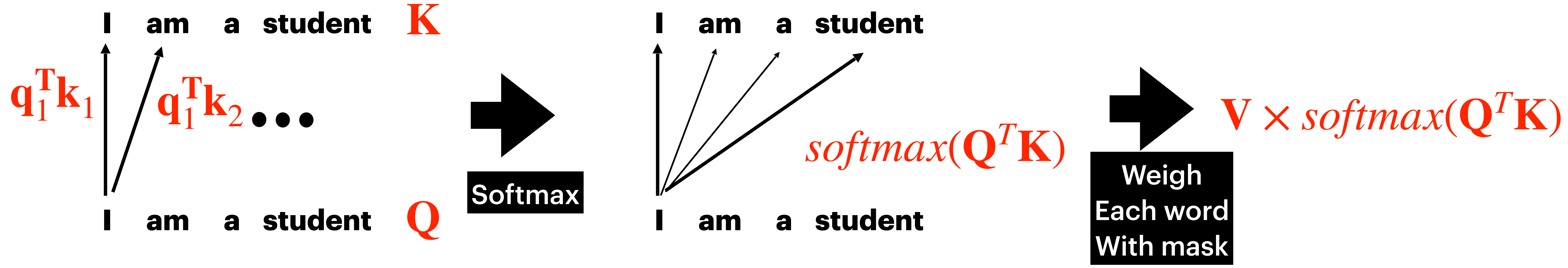
$$\mathbf{K} = \{\mathbf{k}_1, \mathbf{k}_2, \ldots \mathbf{k}_T\}; \mathbf{k}_t \in \mathscr{R}^D$$

I am a student

$$\mathbf{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \ldots \mathbf{q}_T\}; \mathbf{q}_t \in \mathscr{R}^D$$

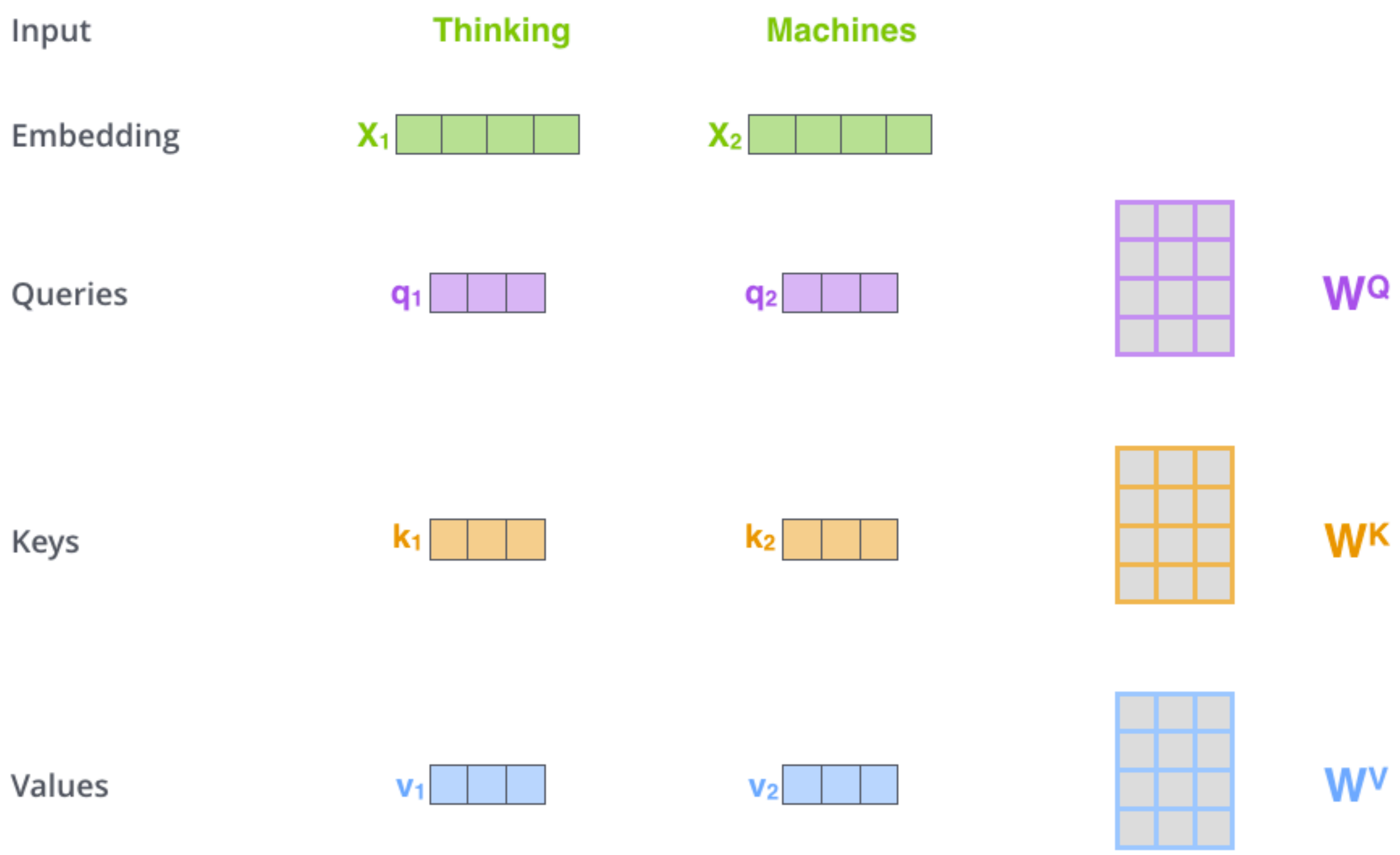
$$\mathbf{K} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_T\}; \mathbf{v}_t \in \mathscr{R}^D$$

I am a student

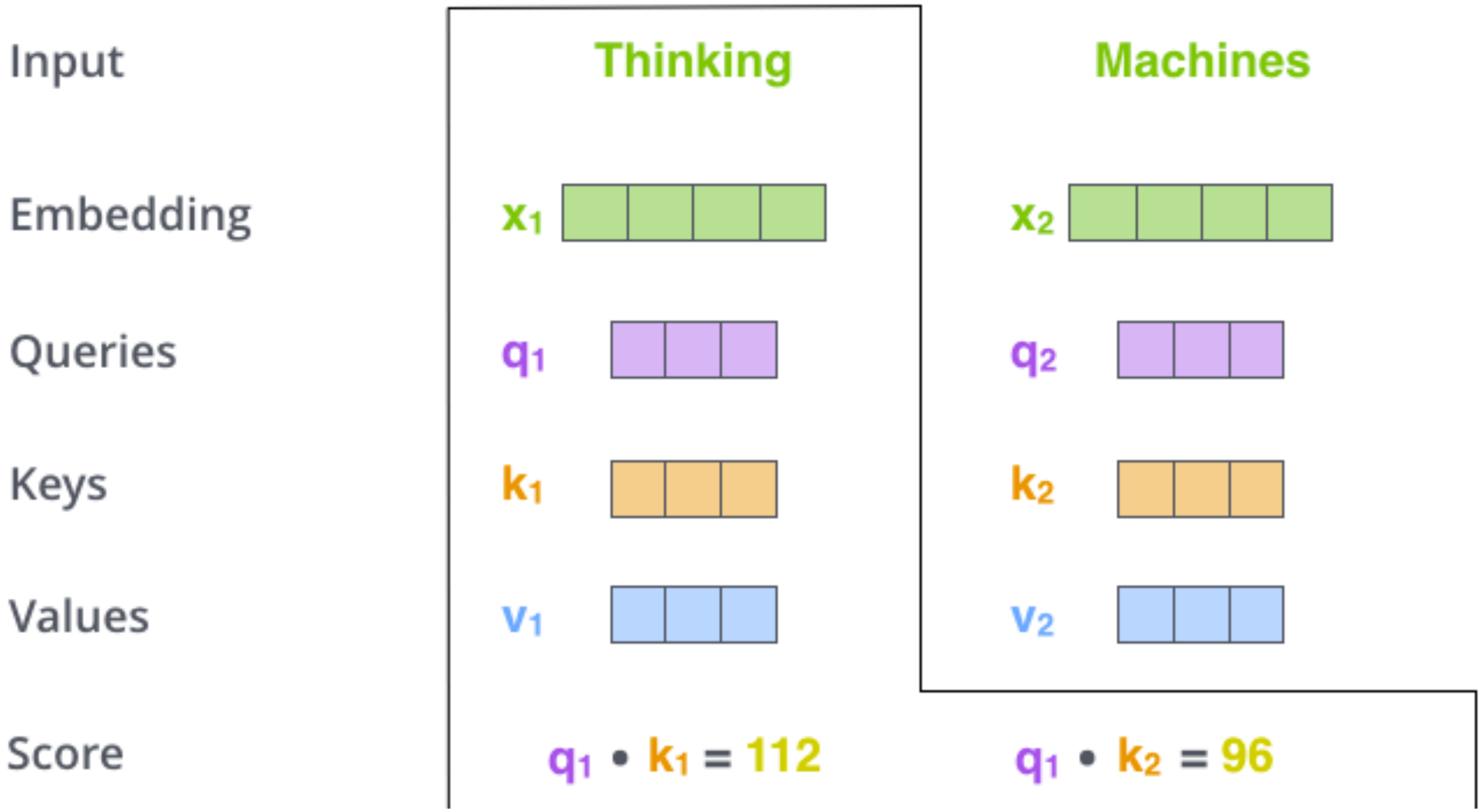


Pics taken from : https://jalammar.github.io/illustrated-transformer/

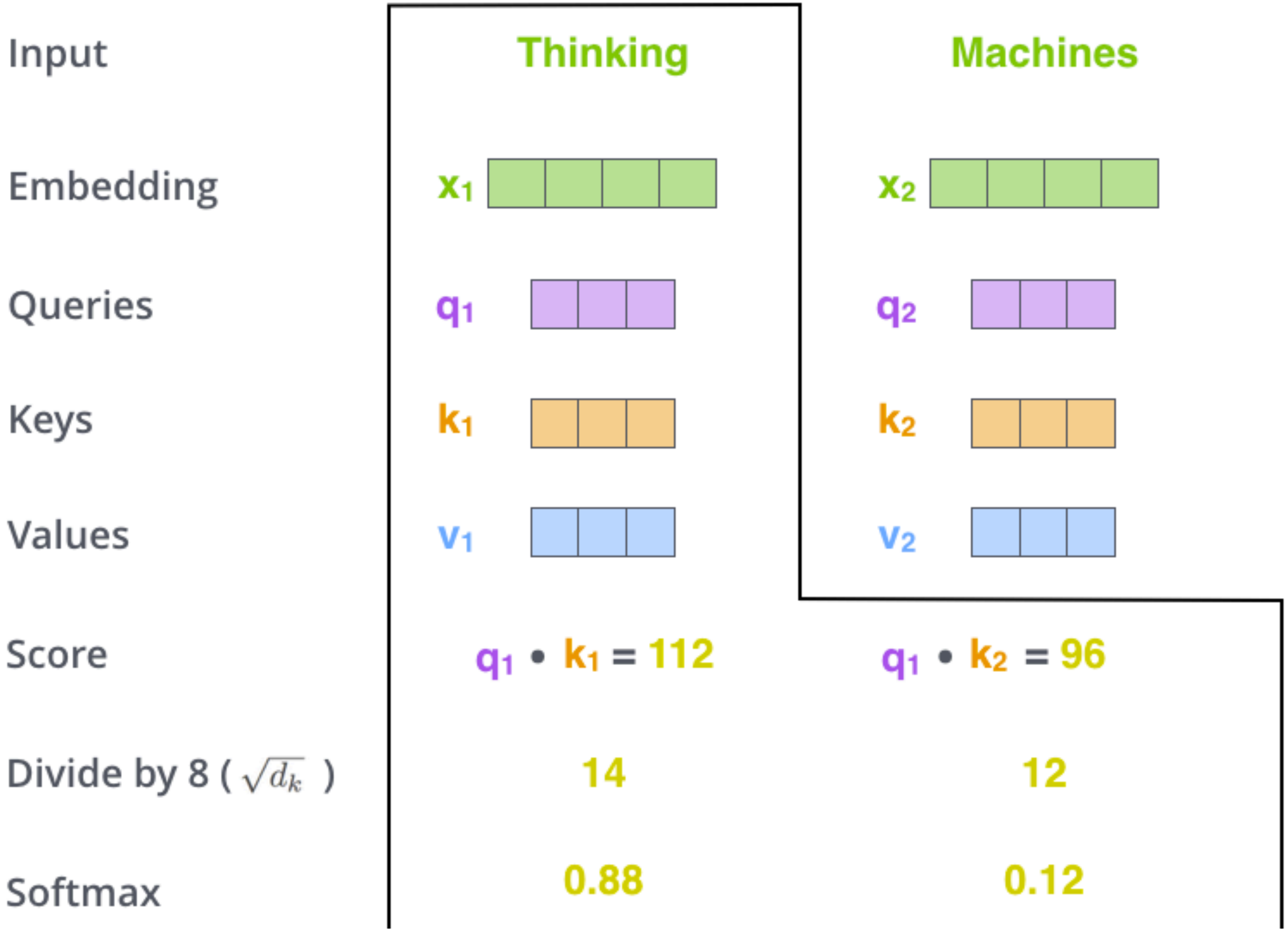
Transformers - self-attention



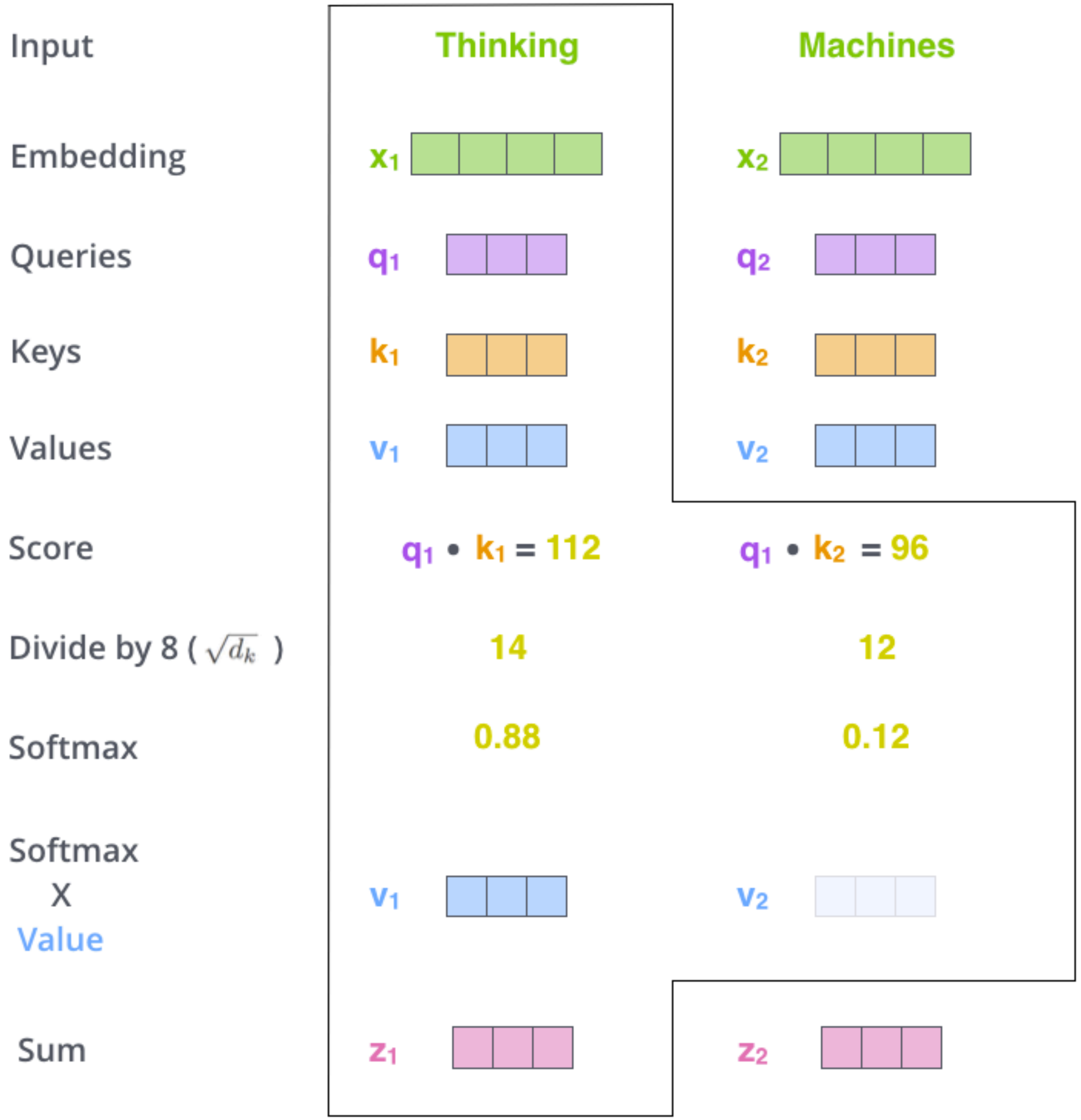
Transformers - self-attention



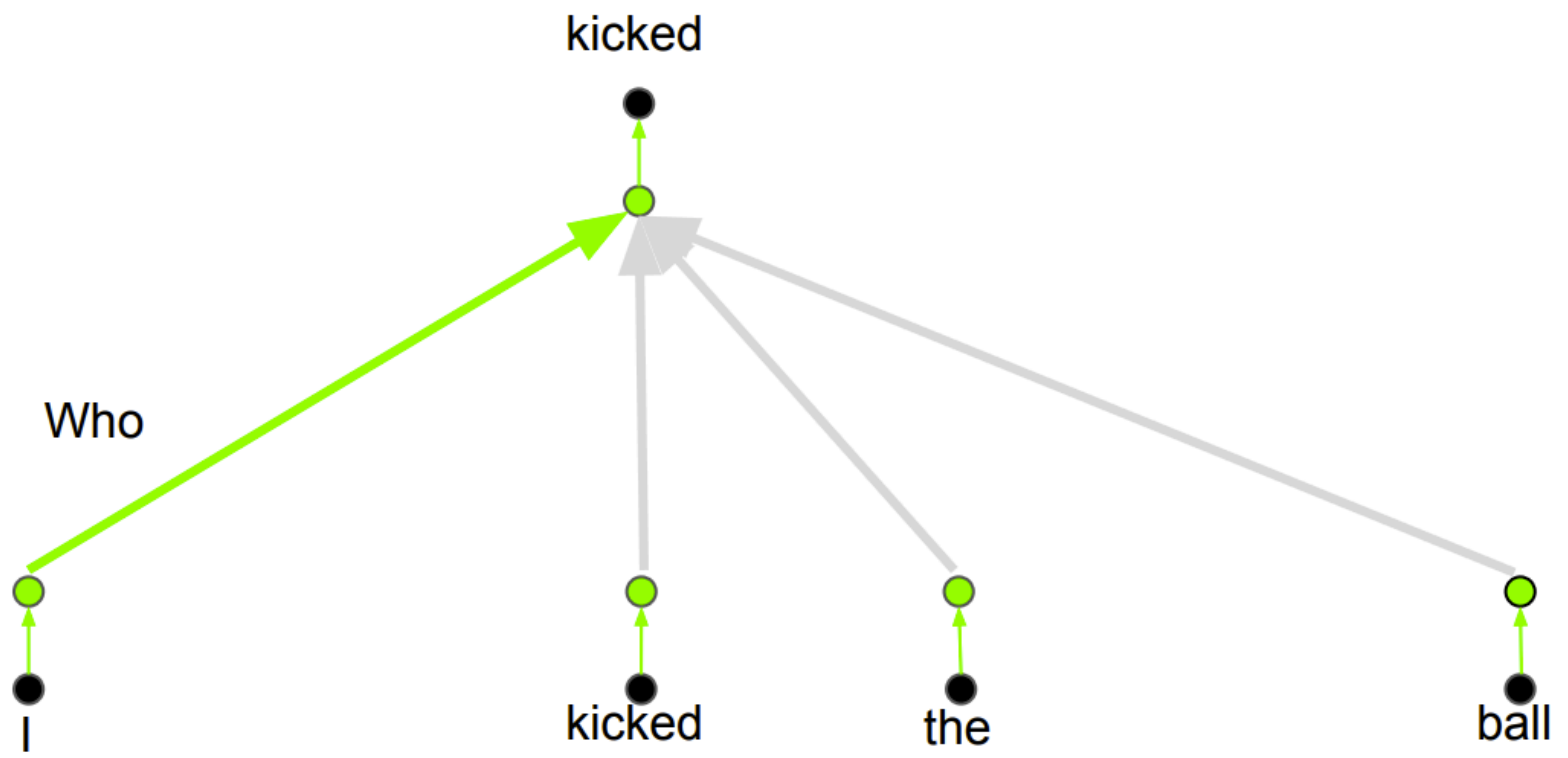
Transformers - self-attention



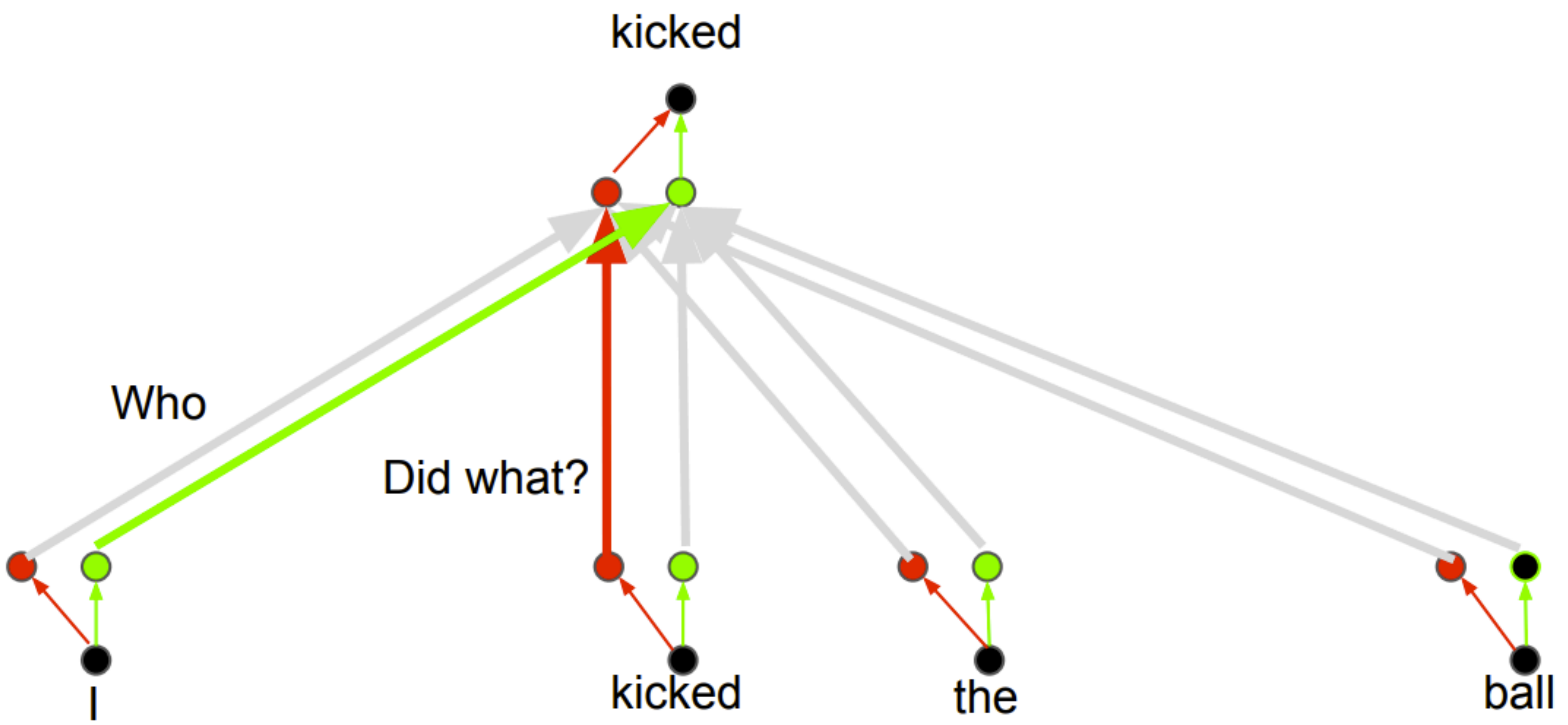
Transformers - self-attention



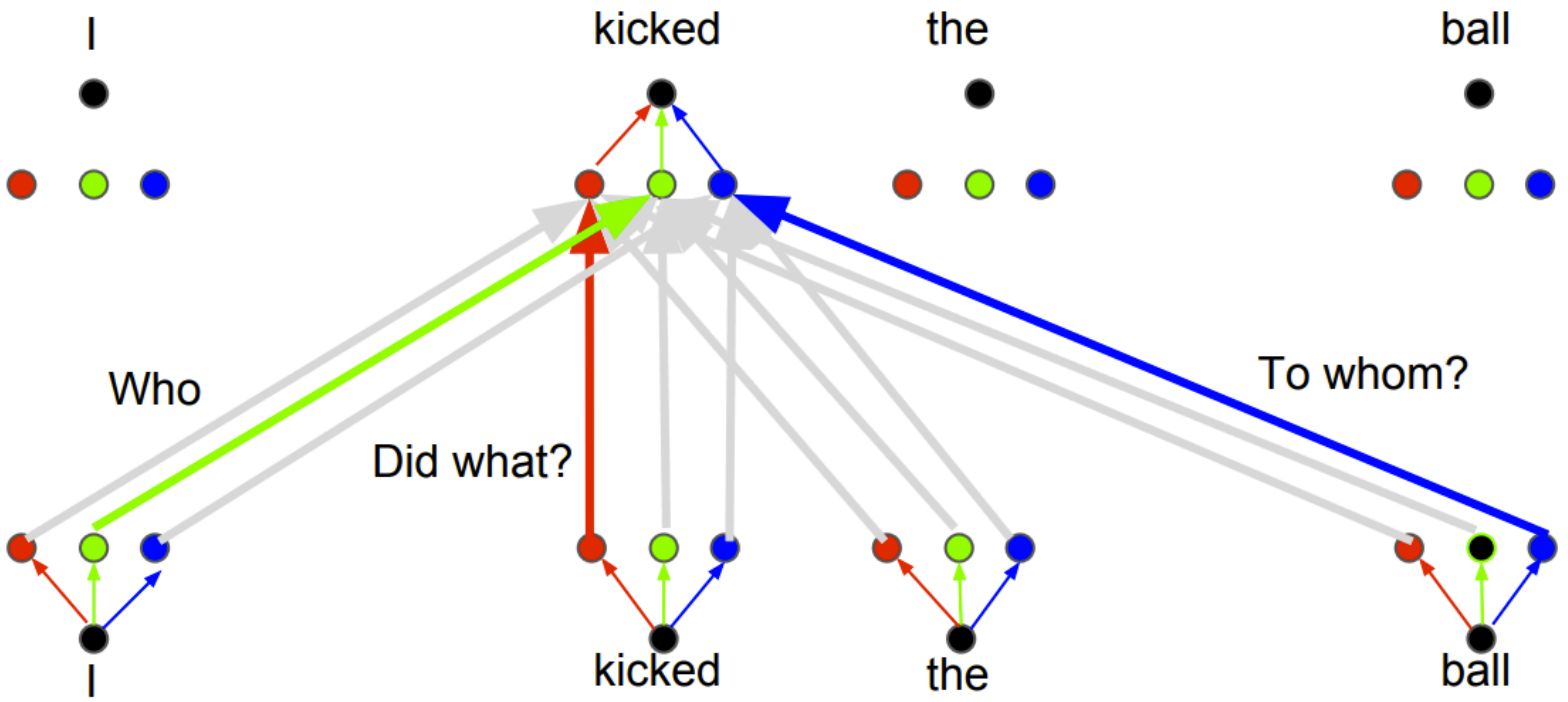
Multi-head



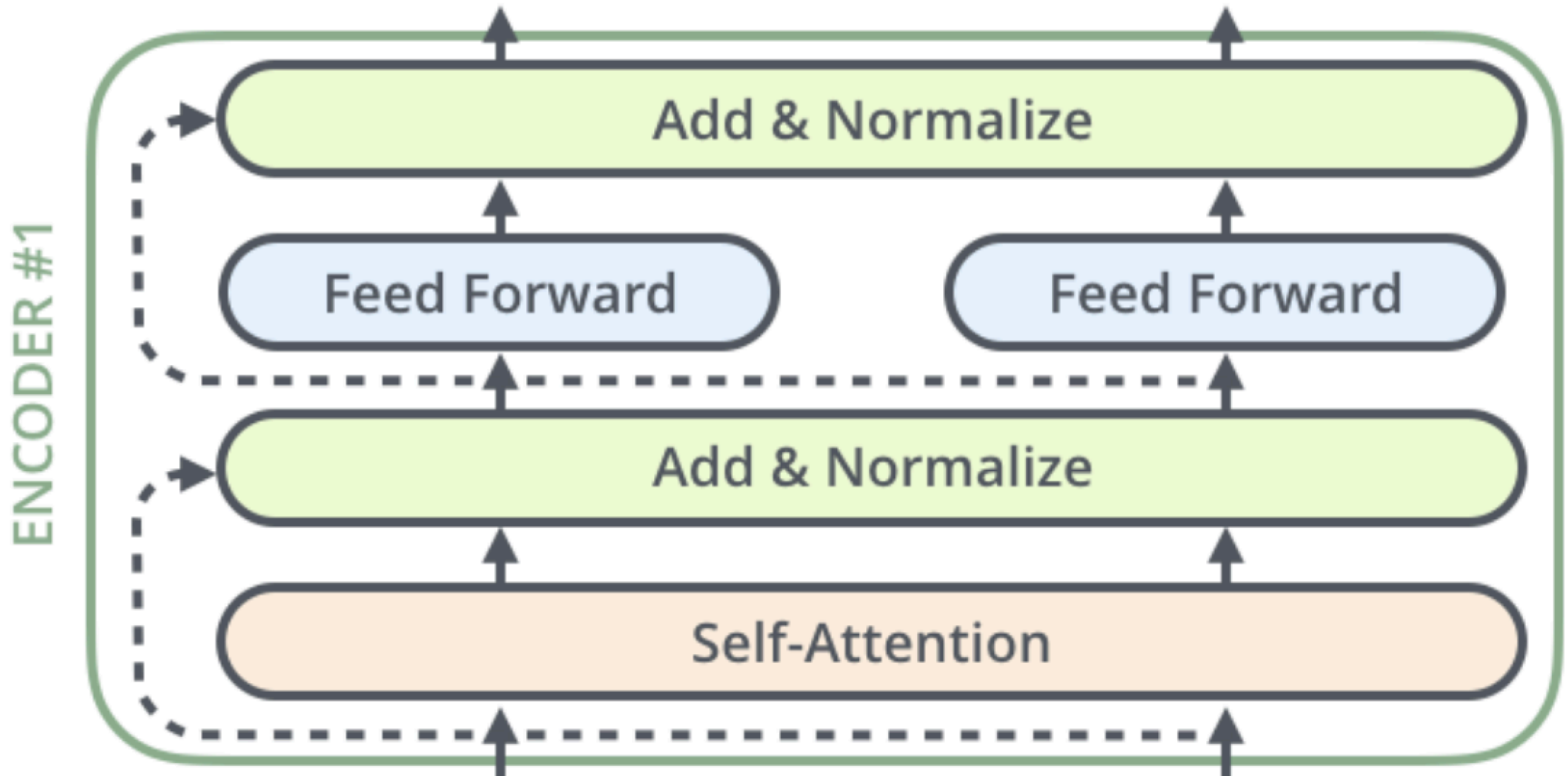
Multi-head



Multi-head

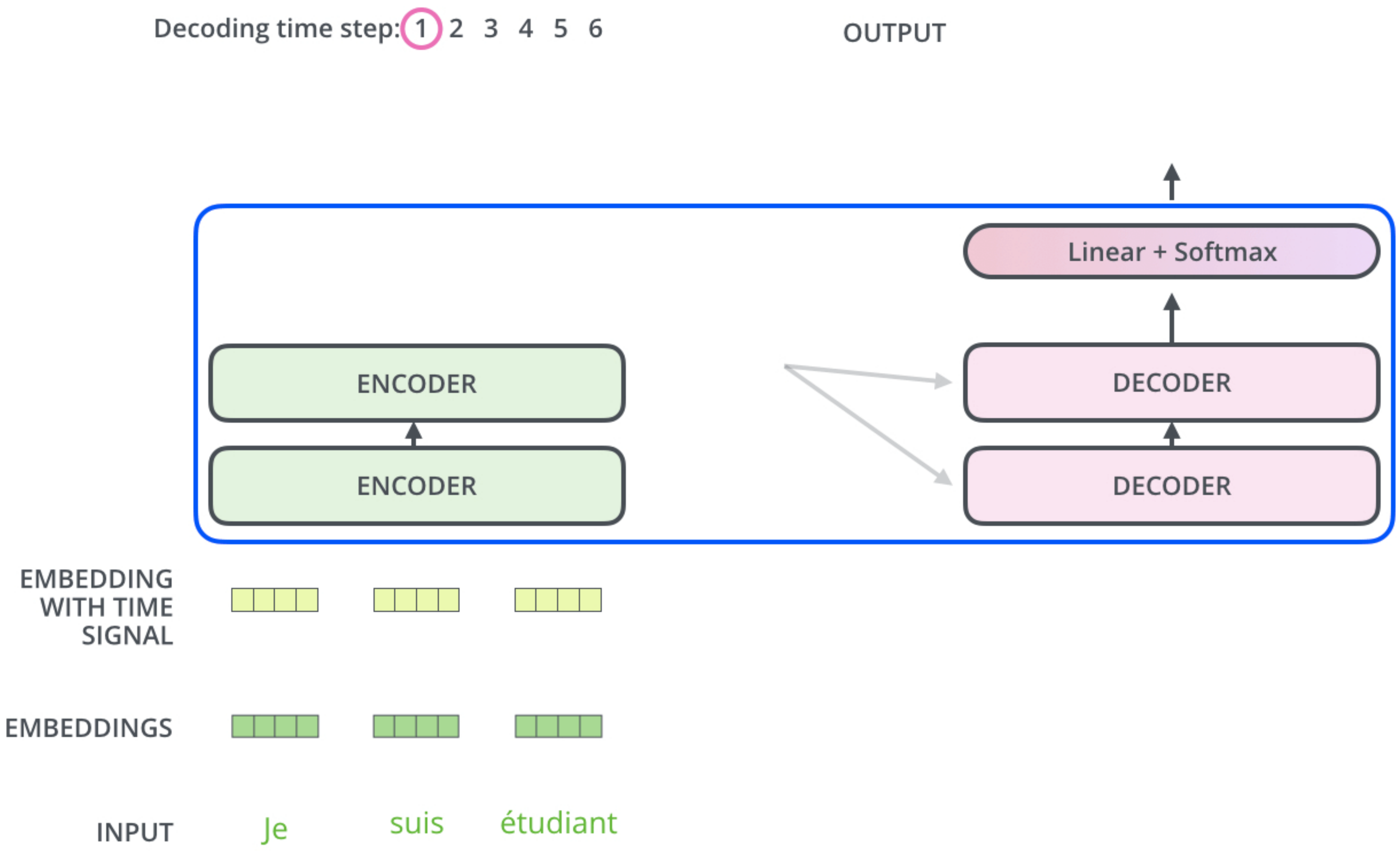


Transformer encoder



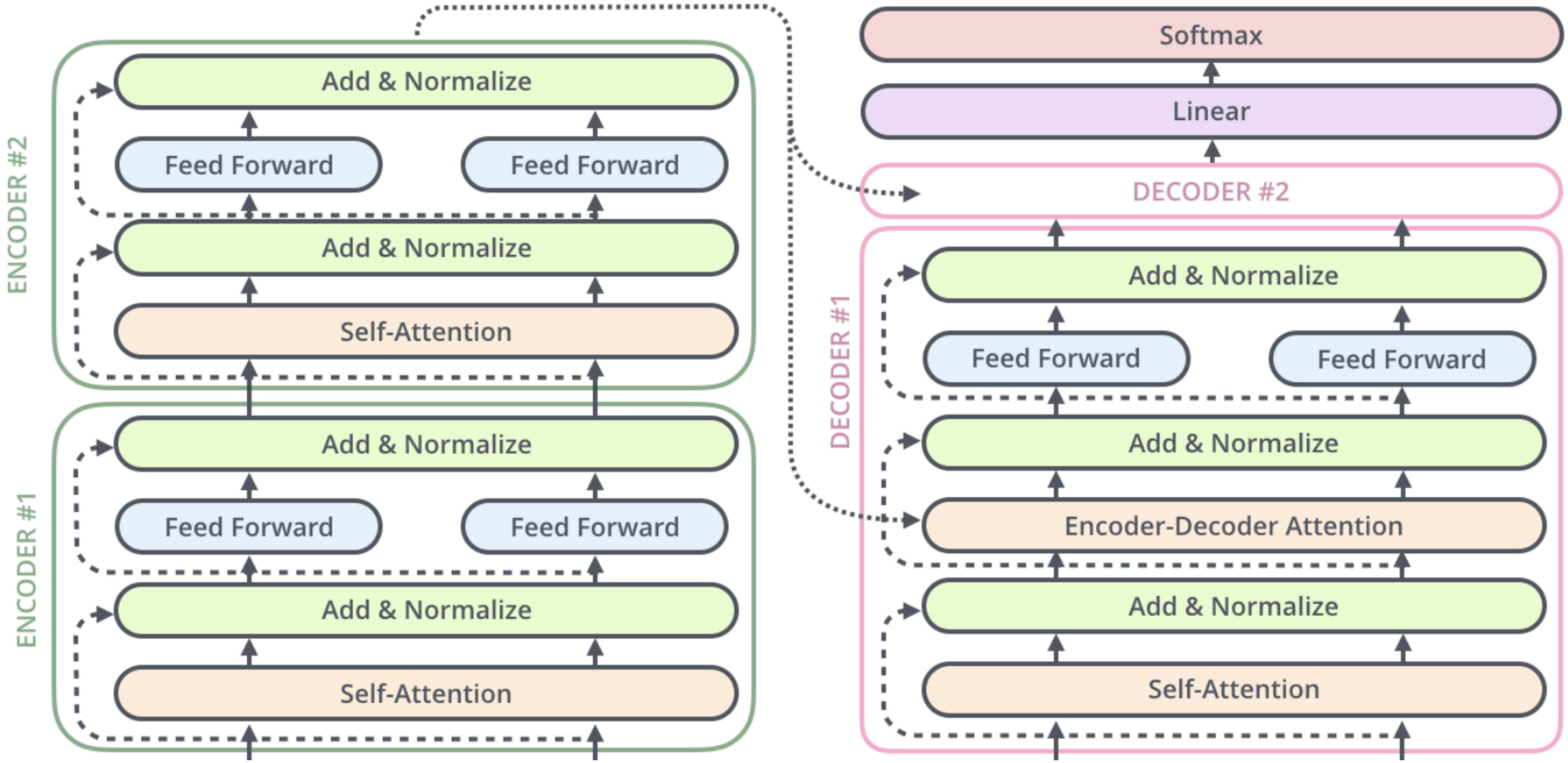
Pics taken from : https://jalammar.github.io/illustrated-transformer/

Transformer



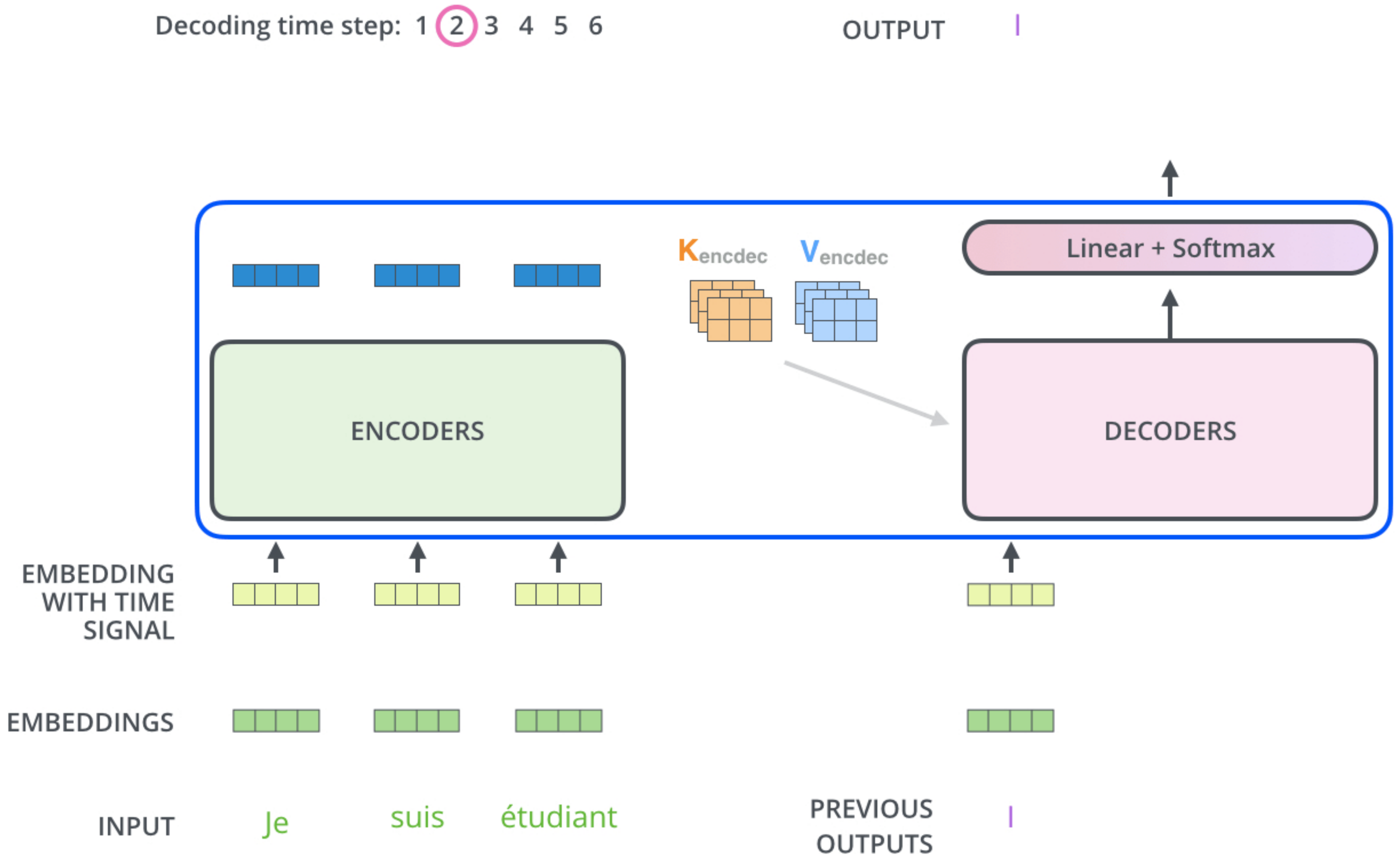
Pics taken from : https://jalammar.github.io/illustrated-transformer/

Transformer decoder



Pics taken from : https://jalammar.github.io/illustrated-transformer/

Transformer Example



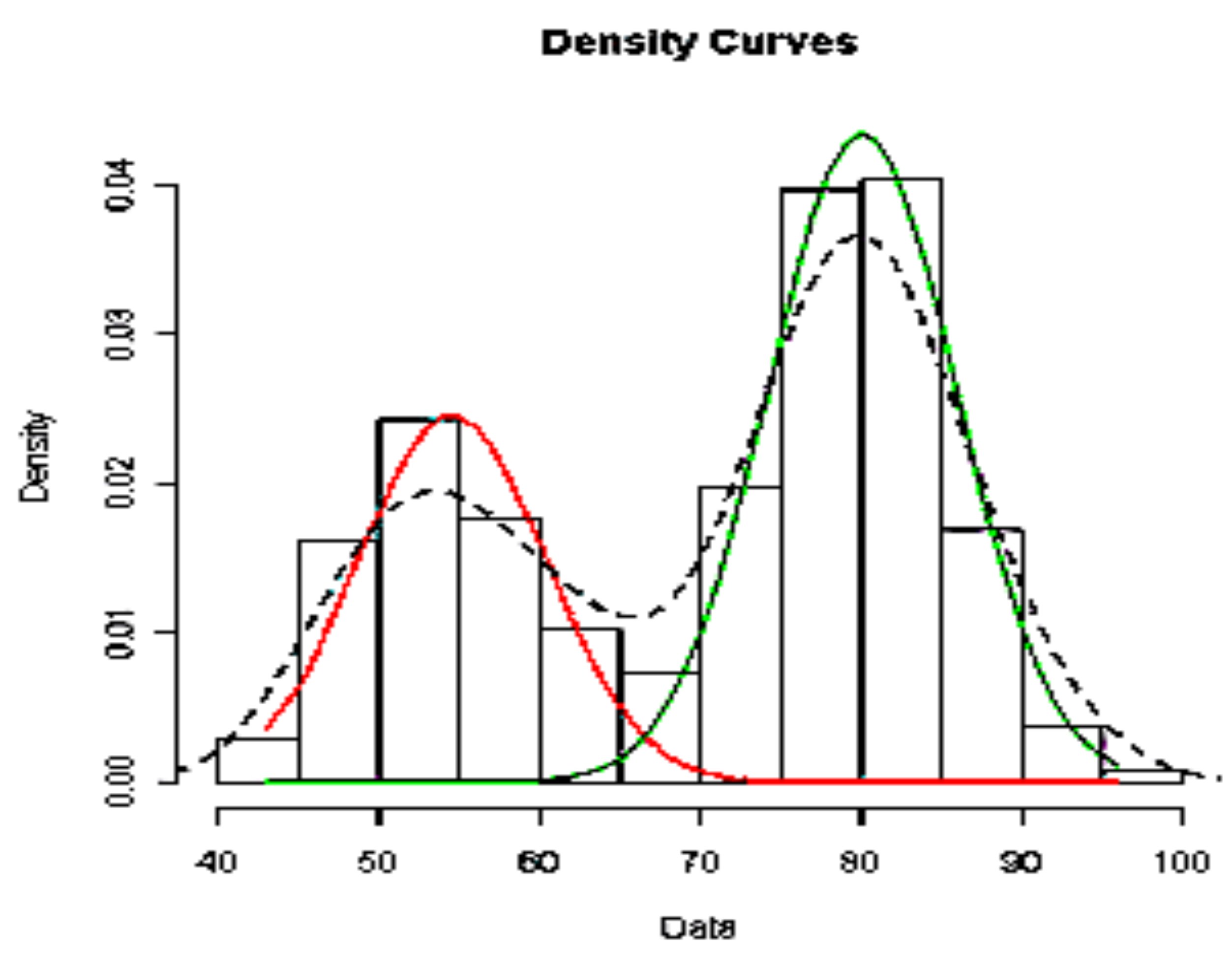
Pics taken from : https://jalammar.github.io/illustrated-transformer/

Unsupervised Learning

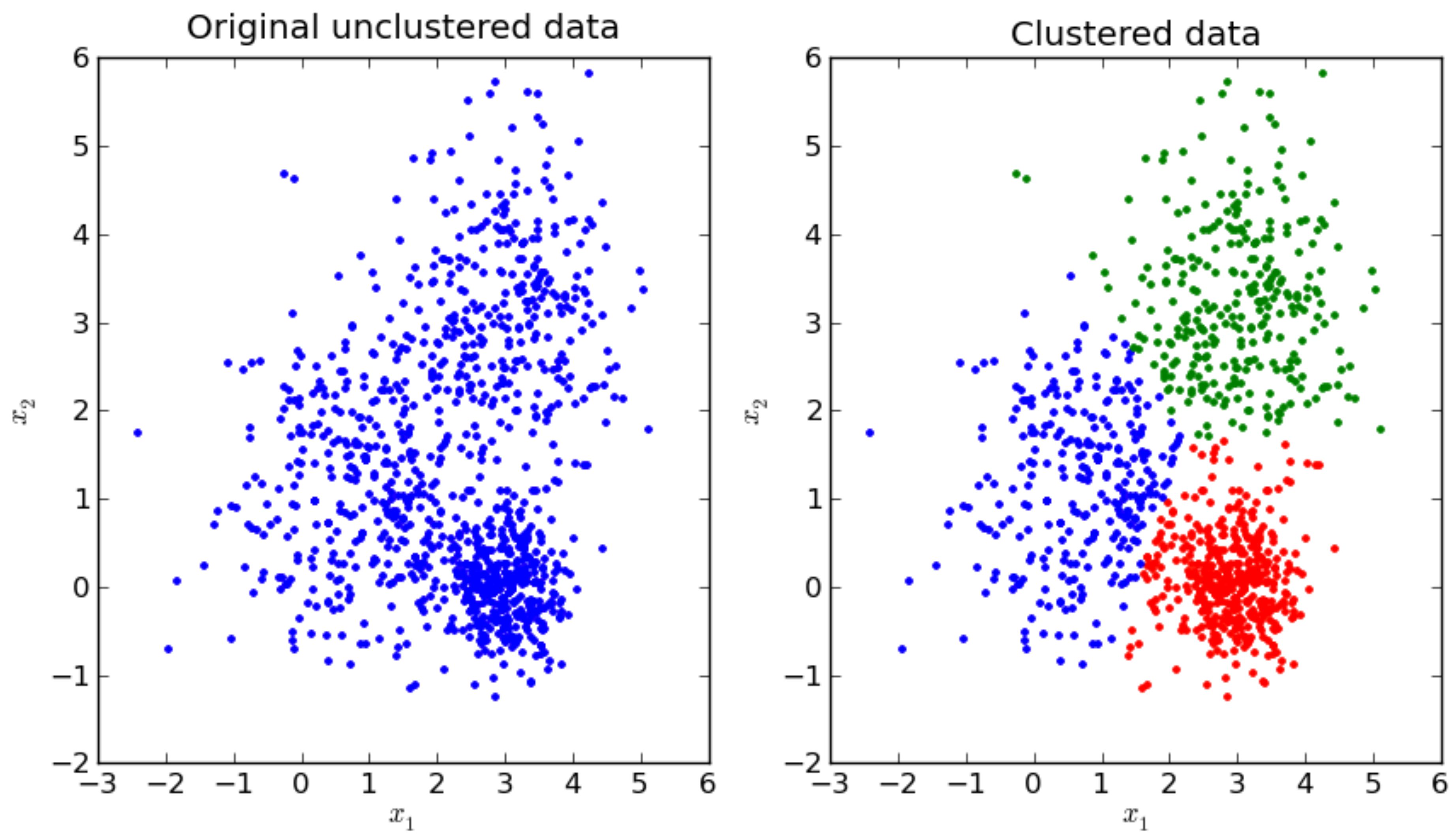
Unsupervised Learning

- Developing models that do not need labels
  - May model the generation of data.
  - May allow generation of new data samples
- Broad strategies for unsupervised learning

Learning the distribution of the data

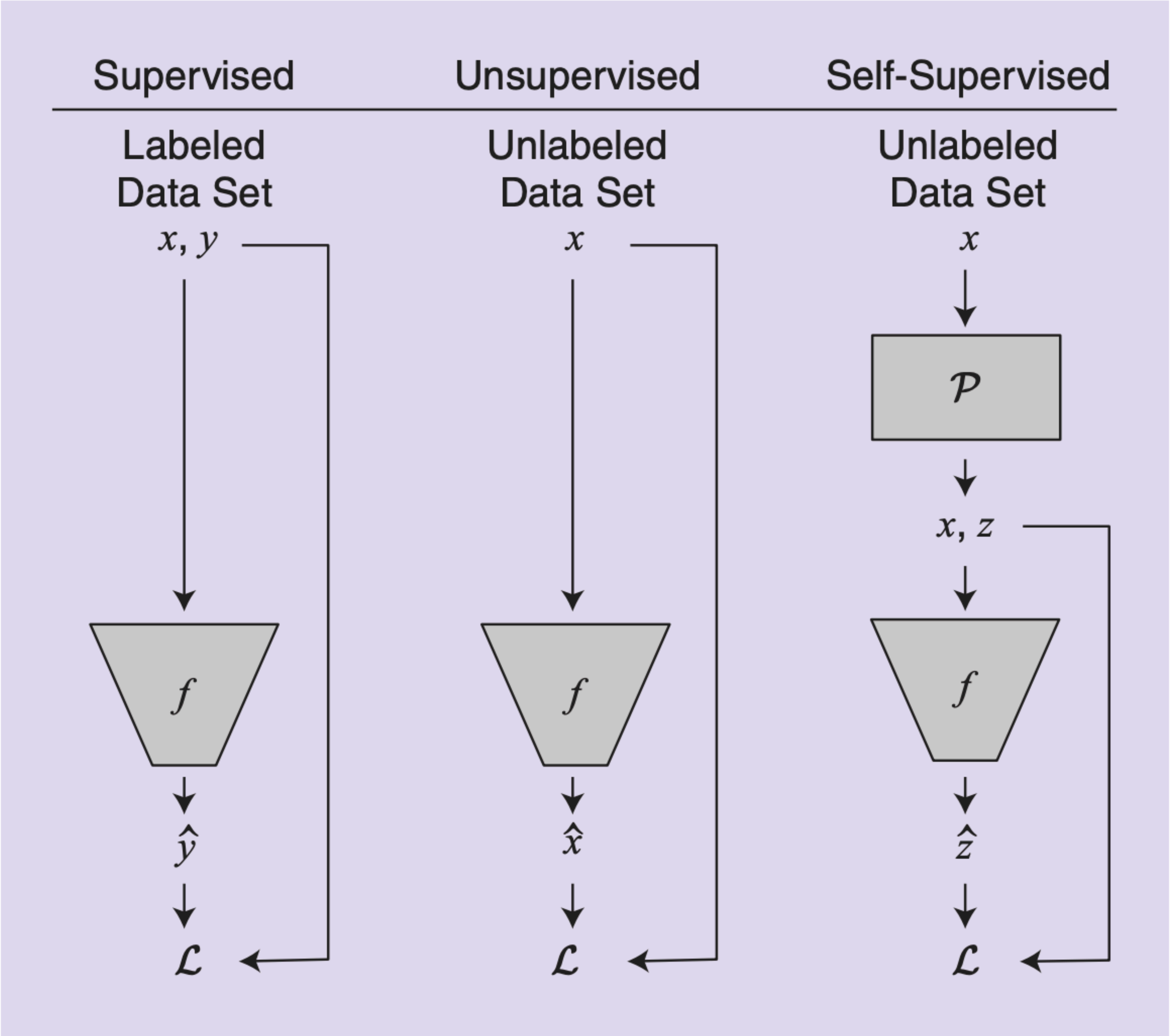


Detecting clusters in the data



Self supervision

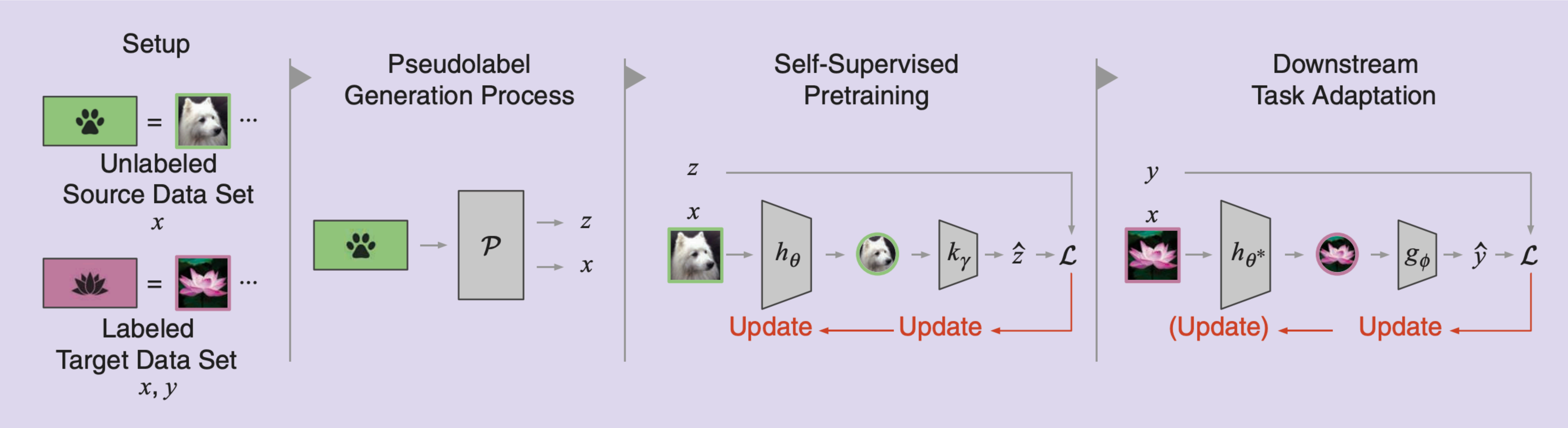
- ✦ Different from supervised and unsupervised learning
  - ✳ Does not perform distribution learning or reconstruction
  - ✳ Uses a pretext task
  - ✳ Performing contrastive or predictive learning
- ✦ Using large volumes of unsupervised data



Ericsson, Linus, et al. "Self-supervised representation learning: Introduction, advances, and challenges." *IEEE Signal Processing Magazine* 39.3 (2022): 42-62.

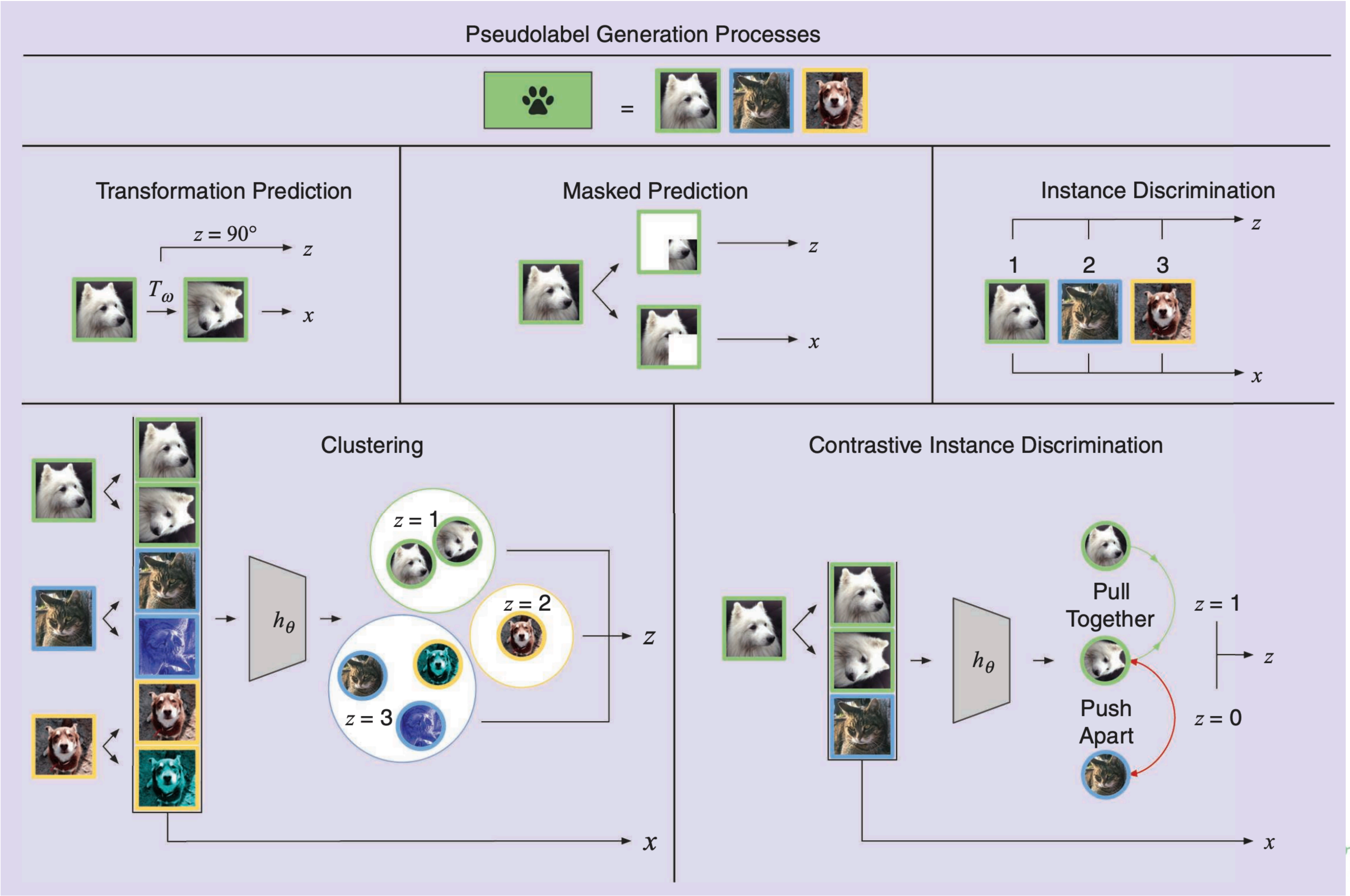
Self supervision - principle

- Two levels of modeling with unsupervised data
  - Generating a pseudo-label
  - Learning the upstream model
- Downstream task performs fine-tuning of the SSL model.



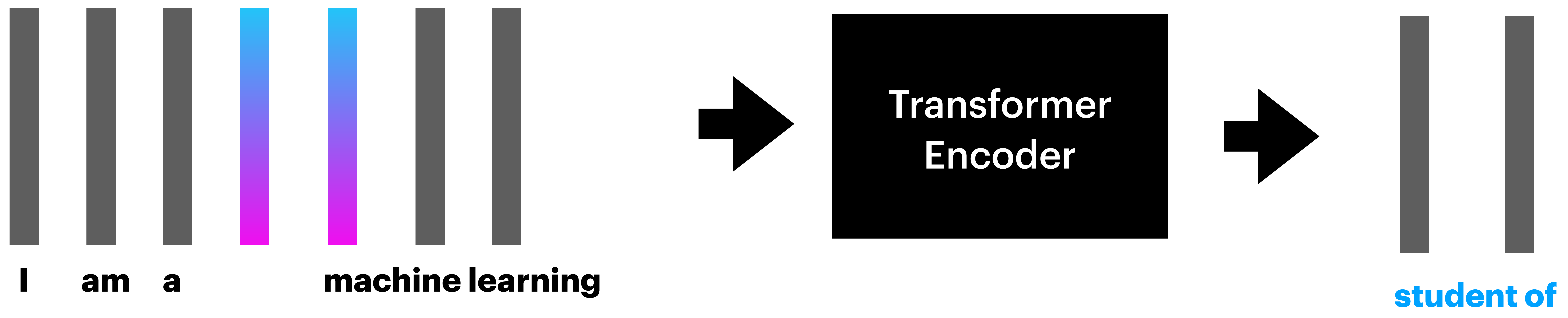
Ericsson, Linus, et al. "Self-supervised representation learning: Introduction, advances, and challenges." *IEEE Signal Processing Magazine* 39.3 (2022): 42-62.

Self supervision - pre-text task



Self-supervision as a task

- Masking out portions of the input data
  - Pass the rest of the embeddings (with zeros or random entries at the masked locations) to the transformer encoder
  - Have the model predict the word tokens in the masked portions - Masked Language Modelling (MLM)



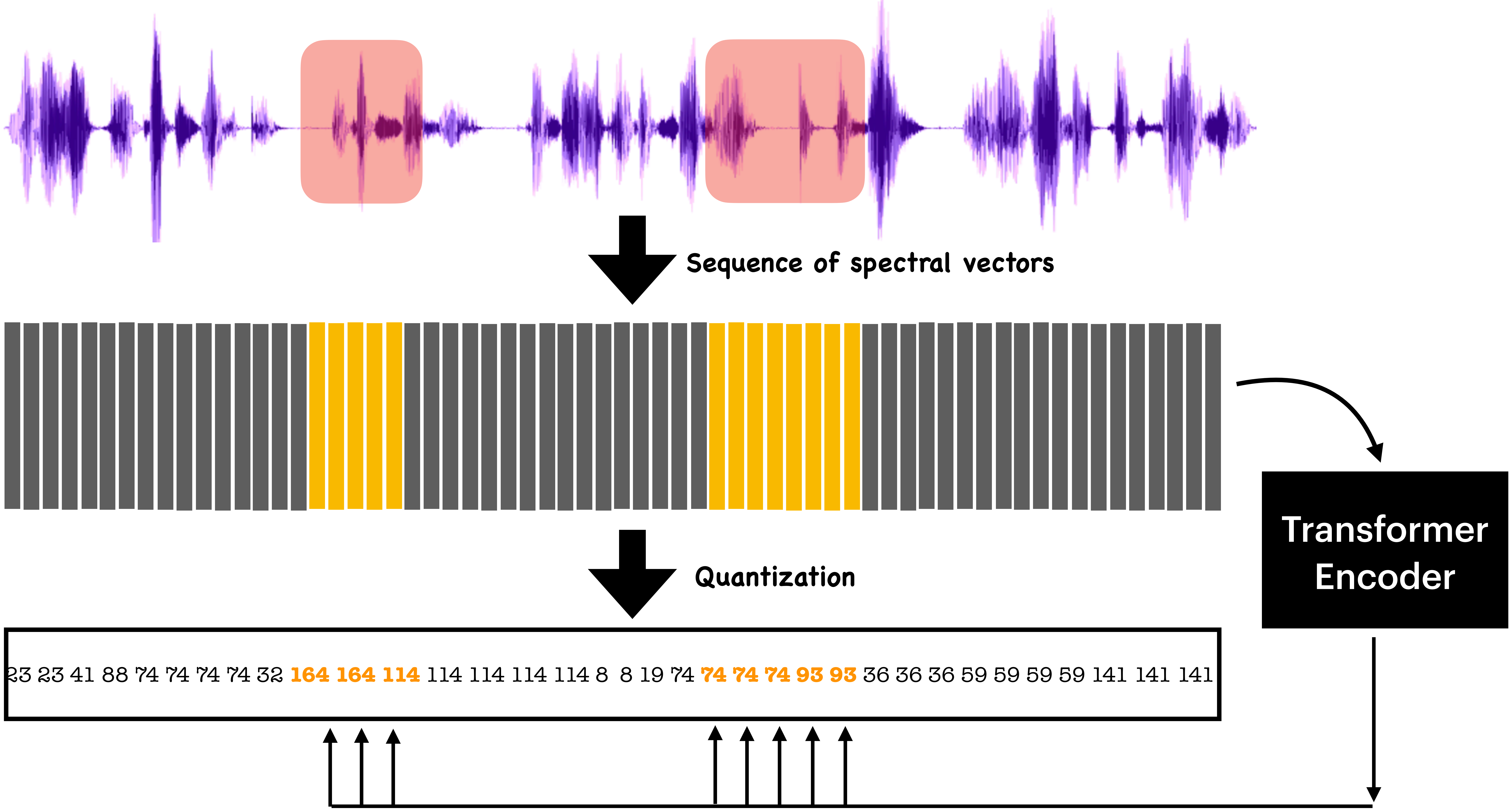
Large language models (LLMs)

- Extending the task of self-supervision
- Mine lots of text data
  - ✳ Crawled from the web, as well as, from other resources.
- Design the model with large capacity (Millions —> Billions of parameters)
- Pre-train the model
  - ✳ With MLM and similar style of losses
  - ✳ High resource of computations.
- Final trained model can be fine-tuned for supervised tasks
  - ✳ Load the parameters as initialization and perform supervised learning.

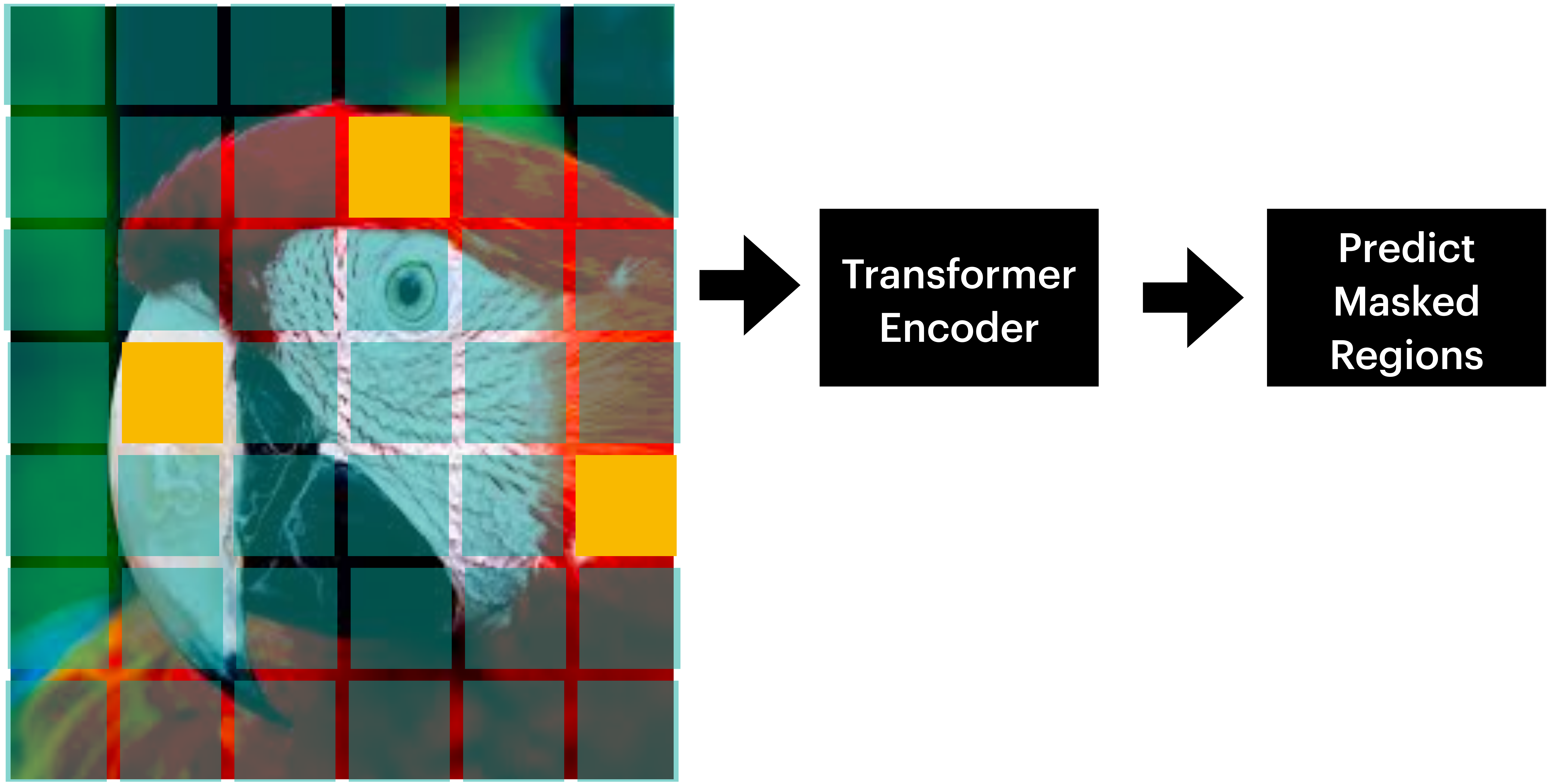
Large language models (LLMs)

- Self-supervised learning
  - ✳ Has shown emergent abilities to generalise to wide variety of downstream tasks.
    - ✓ Tasks that the model was not trained on
    - ✓ Not seen in smaller models
  - ✳ Enables to build reasoning capabilities in the model.
  - ✳ Applicable for several domains - text, speech and images.

Self-supervision in audio - wav2vec



Self-supervision in images - Vision Transformer



Future works (some already underway)

- Multi-modal
  - ✳ Incorporating learning across modalities
    - ✓ Create a domain specific encoder/decoder and learning the joint language model.
- Combining some labeled data with the self-supervised data to further improve the models.
  - ✓ Current models like GPT use human feedback.
- Understanding the risks and vulnerabilities of these models.

THANK YOU

Sriram Ganapathy and TA team
LEAP lab, C328, EE, IISc
sriramg@iisc.ac.in