

NATURAL LANGUAGE PROCESSING

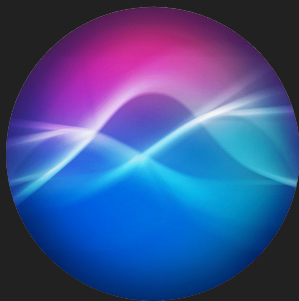


Word embeddings, RNN
et Transformers

Introduction



ChatGPT



DeepL



OpenAI GPT-3

Preprocessing

Après avoir capturé la souris, le chat l'a mangé

→ après avoir capturé souris chat mangé

→ [“après”, “avoir”, “capturé”, “souris”, “chat”, “mangé”]

Stemming :

Étudiant, étudier, études → étud

Université, universitaire, universel → univers ?

Représenter les mots

“Le chat boit le lait”

0	1	2	3
le	chat	boit	lait

→ Lait > Chat ?

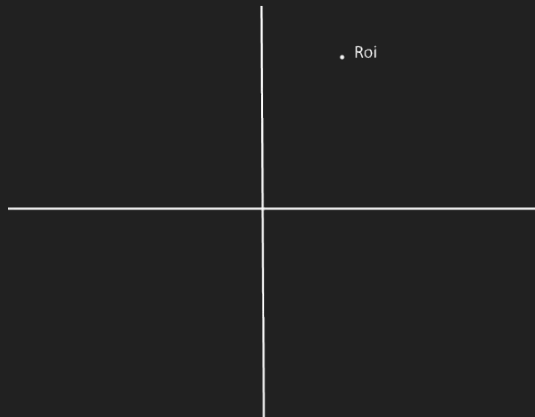
le → [1] → [1, 0] → [1, 0, 0] → [1, 0, 0, 0]

chat → [0, 1] → [0, 1, 0] → [0, 1, 0, 0]

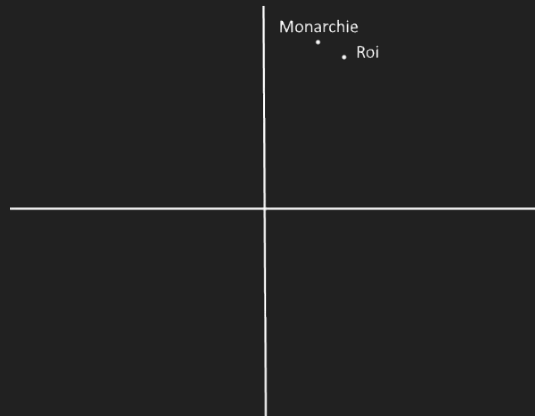
boit → [0, 0, 1] → [0, 0, 1, 0]

lait → [0, 0, 0, 1]

Représenter les mots : Embeddings



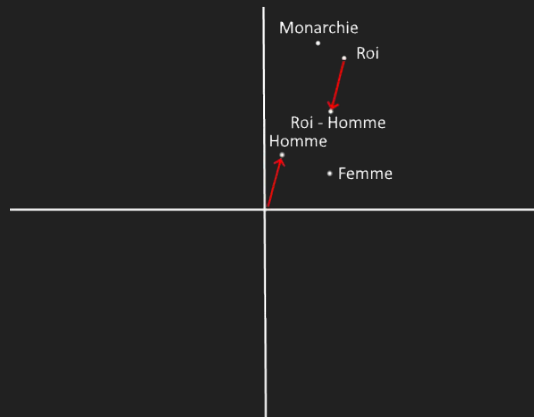
Représenter les mots : Embeddings



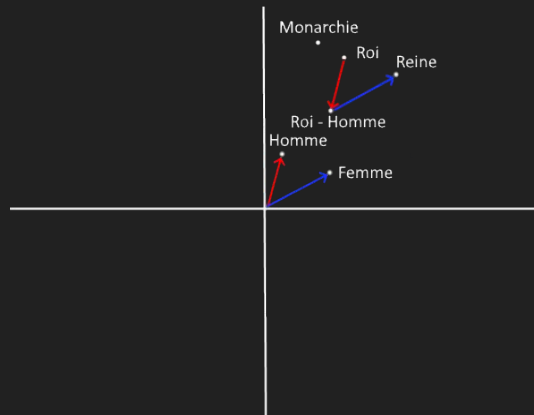
Représenter les mots : Embeddings



Représenter les mots : Embeddings



Représenter les mots : Embeddings



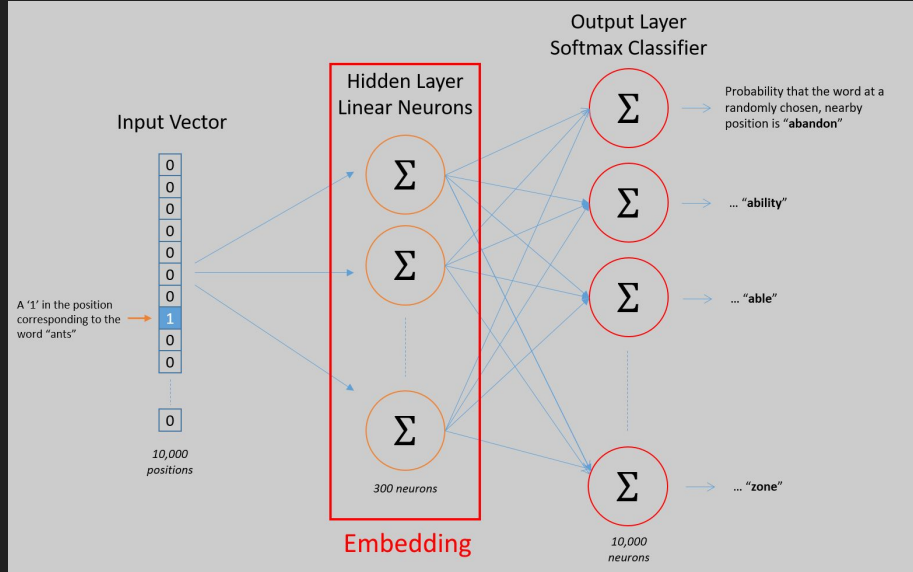
Représentation des mots : Word2Vec

le $\rightarrow [1, 0, 0, 0]$

chat $\rightarrow [0, 1, 0, 0]$

boit $\rightarrow [0, 0, 1, 0]$

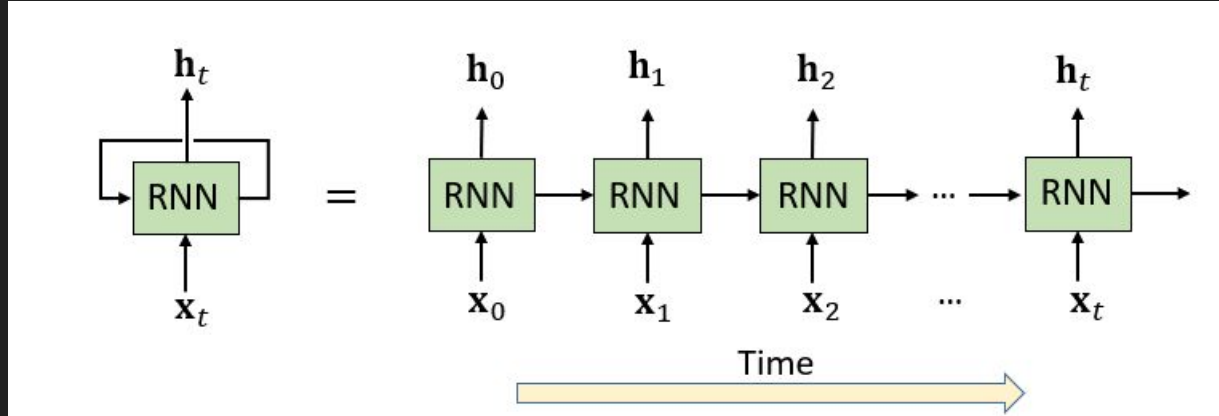
lait $\rightarrow [0, 0, 0, 1]$



Entrée : "le boit" $\rightarrow [[1, 0, 0, 0], [0, 0, 1, 0]]$

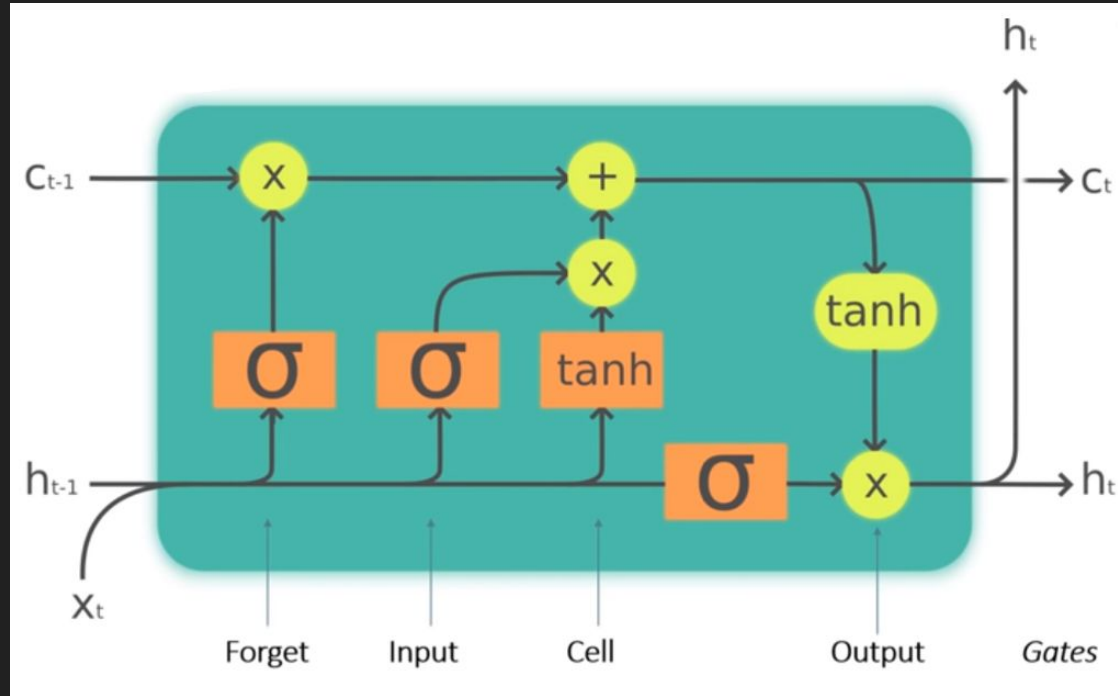
Sortie attendue : $[0, 1, 0, 0]$

Réseaux de neurones récurrents (RNN)



Le **ciel**, malgré une journée grise et nuageuse, avait enfin fini par se dégager et à laisser apercevoir une magnifique couleur ____.

Long Short-Term Memory (LSTM)



Transformers et Attention

