Les transformers sont des systèmes qui ont révolutionné le monde de l'apprentissage de langage notamment en 2017 ou google a sorti un article pour dire que cela était plus efficace que tout les autres par la suite Google a sorti son llm puis Openai en 2018.

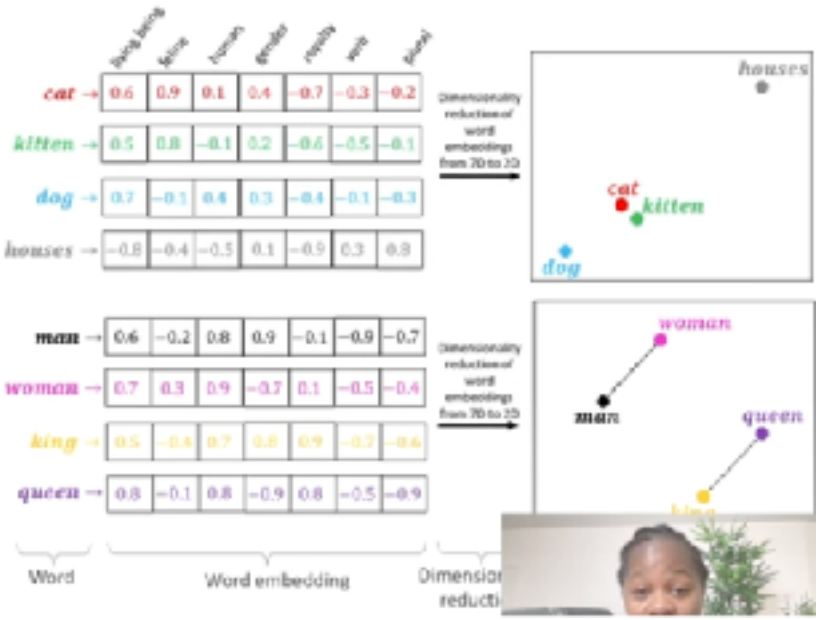
Tokenisation

Cela est un processus de découpage d'un texte en morceaux plus petits comme des mots des lettres ou des phrases.

En deep learning il faut etre capable de transformé ces mots représentation vectorielle cela s 'appelle l'Embedding

Embedding

L'embedding est une représentation vectorielle qui capture la sémantique des mots ou des phrases dans un espace de haute dimension



Comme ici chaque moi aura une représentation vectorielle en 7D qui sera ensuite représenté en 2D comme sur le graphique

Chaque dimension représente une catégorie par exemple la partie humaine le genre etc …

Cela va pouvoir permettre de manipulé ces mots en faisant des calculs dessus etc …

NLP

En deep learning il y a 2 parties l'encodage et le décodage

Encodeur

L’encodeur est l’input c’est a dire qu’il va récupéré ce qu’il y a en entrée pour le transformé et pouvoir le projeté dans un espace par exemple transformé la phrase en la séparant en mot puis transformé les mots en vecteurs pour pouvoir ensuite les manipulé.

Ca peut etre de la convolution CNN comme vue dans l’algorithme de convolution des images cela permettait de transformé les images avec différent filtre donc ca créé par exemple 4 images avec 4 image différentes qui mettent en avant par exemple les bordures ou les textes … ca peut également modifié le padding ou le cadrage de l’image. Puis ensuite il y a l’effet filtre qui efface les les pixels négatif les transforme en 0 etc … Puis enfin le dernier filtres qui permet de réduire la taille de l’image en combinant des pixels ensemble ca permet de réduire par 2 les pixels et d’optimisé le temps d’apprentissage.

Décodeur

Le décodeur son rôle est de récupéré tous ca pour faire la sortie en utilisant un model pour entrainer l’ia

Transformers

Les transformers par rapport aux ancienne technique est beaucoup plus efficace notamment grâce a la parralélisation. Il va par exemple traiter la phrase en entière et non traiter chaque mot 1 par 1.
Cela se fait grâce a différent mécanisme.

Mécanisme d’attention

Le mécanisme d'attention est une technique permettant aux réseaux de neurones de se concentrer sur une partie spécifique de la séquence d’entrée. C’est a dire qu’elle va donner plus d’attention a certaine mot de la phrase grâce a une formule traduite sous python.

Multi-head Attention

Une méthode qui utilise plusieurs "têtes" d'attention en parallèle pour capturer diverses relations dans les données.

Cela va donc divisé les vecteurs par le nombre de tête puis refaire le mécanisme d’attention c’est comme un système parallèle au mécanisme d’attention.