

# Les Maths derrière la Génération de Texte en IA

Dr. Clotilde Djuikem

## Problématique

Comment une intelligence artificielle génère-t-elle des réponses en utilisant des mathématiques ?

## Processus en simplifié 6 étapes

- Encodage du texte avec l'algèbre linéaire.
- Analyse contextuelle avec les probabilités.
- Passage à travers un réseau neuronal (Transformers).
- Optimisation par descente de gradient.
- Évaluation par la théorie de l'information.
- Restitution de la réponse finale.

## Transformation des mots en vecteurs

Chaque mot est transformé en un vecteur dans un espace de dimension  $d$  :  $\mathbf{w} = (w_1, w_2, \dots, w_d) \in \mathbb{R}^d$

Exemple pour le mot "Fourier" :

$$\mathbf{w}_{\text{Fourier}} = (-0.12, 0.54, \dots, -0.89)$$

L'ensemble du texte est représenté par une matrice :

$$\mathbf{X} = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,d} \\ w_{2,1} & w_{2,2} & \dots & w_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,1} & w_{n,2} & \dots & w_{n,d} \end{bmatrix} \in \mathbb{R}^{n \times d}$$

## Prédiction des mots suivants

L'IA estime la probabilité du mot suivant avec :

$$P(w_{i+1} | w_1, w_2, \dots, w_i) = \frac{\exp(s(w_i, w_{i+1}))}{\sum_j \exp(s(w_i, w_j))}$$

où la similarité cosinus est donnée par :

$$s(w_i, w_j) = \frac{\mathbf{w}_i \cdot \mathbf{w}_j}{\|\mathbf{w}_i\| \|\mathbf{w}_j\|}$$

## Mécanisme d'auto-attention

Les modèles Transformers utilisent l'auto-attention :

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V$$

où :

- $Q, K, V$  sont les matrices de requêtes, clés et valeurs,
- $d_k$  est la dimension du vecteur clé.

## Ajustement des poids

Pour minimiser l'erreur, la descente de gradient ajuste les paramètres :

$$W^{(t+1)} = W^{(t)} - \eta \nabla_W \mathcal{L}(W)$$

où :

- $\eta$  est le taux d'apprentissage,
- $\nabla_W \mathcal{L}(W)$  est le gradient de la fonction de perte.

## Entropie de Shannon

L'IA mesure la diversité de la réponse avec :

$$H(X) = - \sum_i P(x_i) \log P(x_i)$$

### Interprétation :

- Si  $H(X)$  est faible, la réponse est trop répétitive.
- Si  $H(X)$  est trop élevée, la réponse est aléatoire.

## Restitution Optimisée de la Réponse

Après toutes ces étapes, l'IA génère la réponse finale optimisée :

### Exemple de Réponse

La transformation de Fourier est une technique permettant de convertir un signal du domaine temporel vers le domaine fréquentiel, ce qui est crucial en IA pour analyser des patterns cachés dans les données.

*Ps: ça c'est 'est une réponse de ChatGPT, en essayant un code avec Python on peut avoir une réponse qui est moins optimisée, et surtout avec des mots qui peuvent être mal placé. C'est pour cette raison qu'il faut améliorer les modèles.*



## Conclusion

Les mathématiques sont représentent du fonctionnement des IA génératives — *Les maths c'est la vie.*

Grâce à l'algèbre linéaire, les probabilités, les Transformers, l'optimisation et la théorie de l'information, une IA est capable de comprendre et de générer du texte de manière cohérente.

Si vous avez trouvé cette application des maths utile et intéressante, n'oubliez pas de liker et de vous abonner à **Clotilde Djuikem** sur LinkedIn et **Tioh Academy** sur Youtube !