



Intelligence

Artificielle & Langage

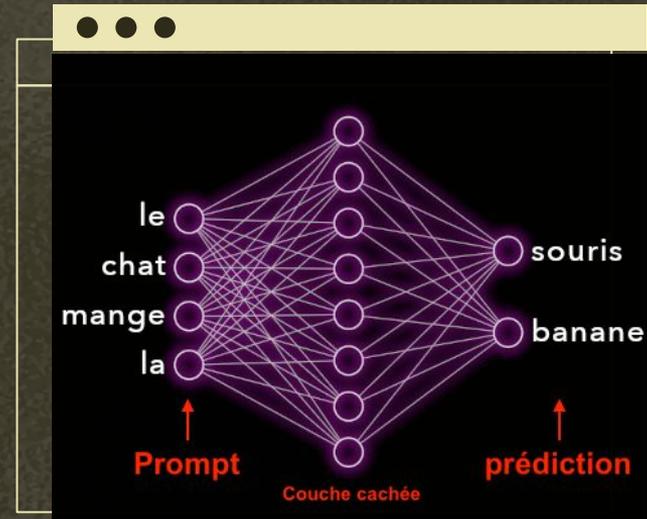
Laurent.Vanni@univ-cotedazur.fr



CNRS - Université Côte d'Azur
UMR 7320 : Bases, Corpus, Langage

Textes et I.A. Définition d'une tâche

En entrée : des mots
En sortie : des mots
Au milieu : la magie des couches cachées.
Représentations abstraites des textes.



Boîte noire ?



L'intelligence artificielle est-elle interprétable ?
explicable ? intelligible ?



Quelles données à interpréter ?

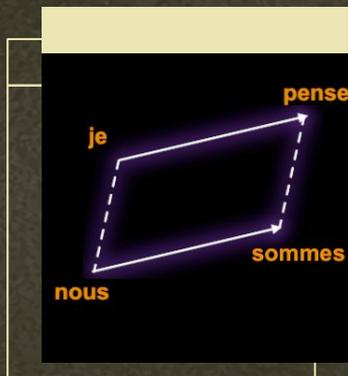
Corpus ou Big Data ?

L'entraînement supervisé :
la force et
la faiblesse
du deep learning



Architectures ? Quelles stratégies ?

Embeddings



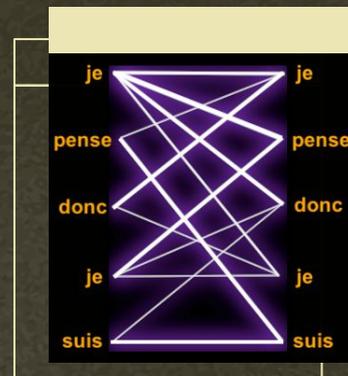
Les mots comme des vecteurs

(Dé)Convolution

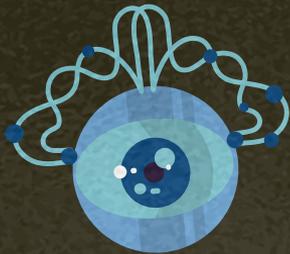


Les mots dans leur co(n)texte

Self-Attention

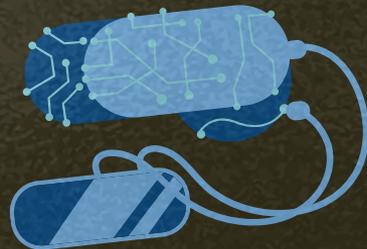


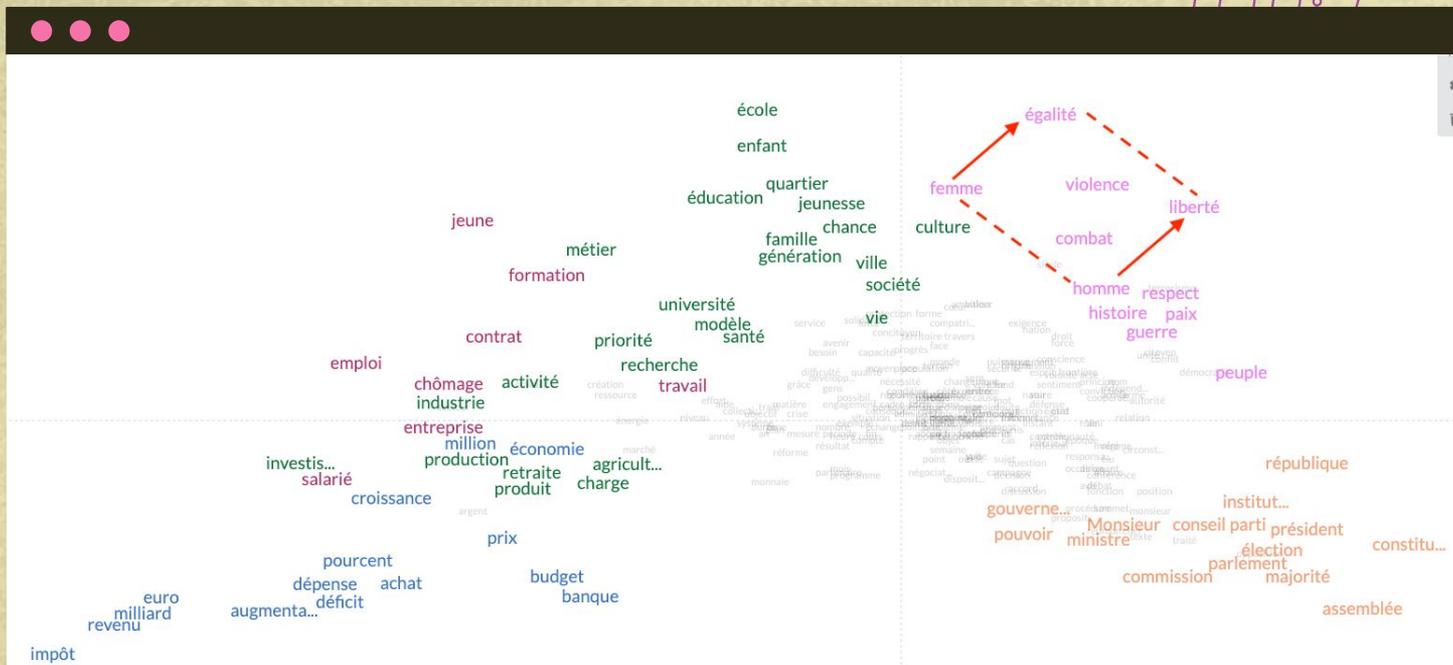
Les associations privilégiées de mots



Embedding

Vectorisation automatique des mots. **Autorise le modèle à utiliser des représentations à n-dimensions** et à faire des calculs vectoriels.



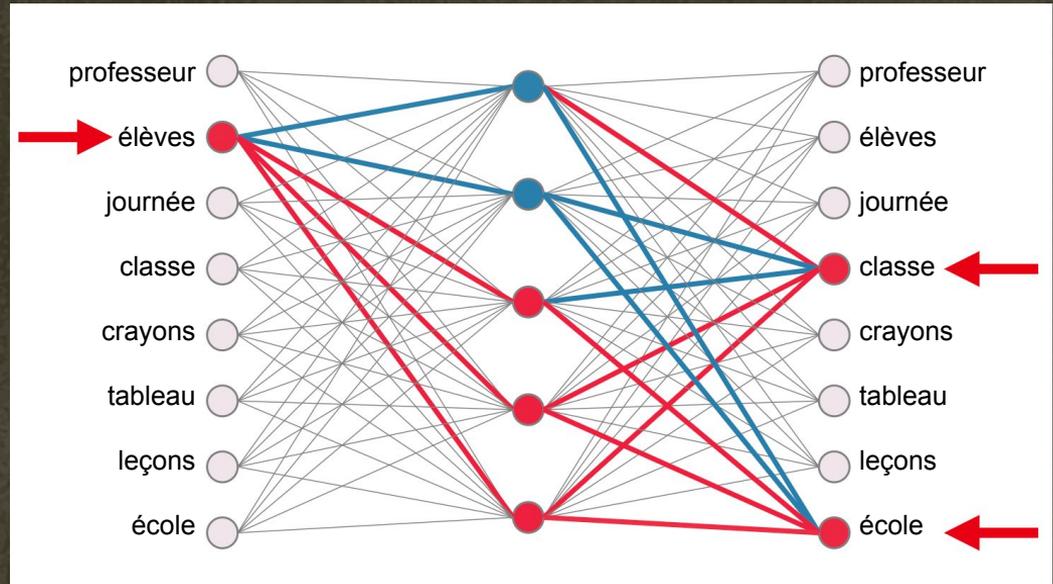


Exemple : l'embedding en politique (discours présidentiels français - Vème république)



Sous-tâche : Représenter des mots

L'école est un lieu animé où chaque matin, les élèves entrent dans leur classe, impatientes de commencer la journée. Ils s'assoient, sortent leurs crayons, et regardent le tableau blanc où le professeur inscrit les leçons du jour. Les élèves, attentifs, posent des questions et prennent des notes, transformant le tableau en un espace de découverte. Chaque crayon tracé marque un pas de plus dans leur apprentissage, guidé par la passion et la sagesse de leur professeur.



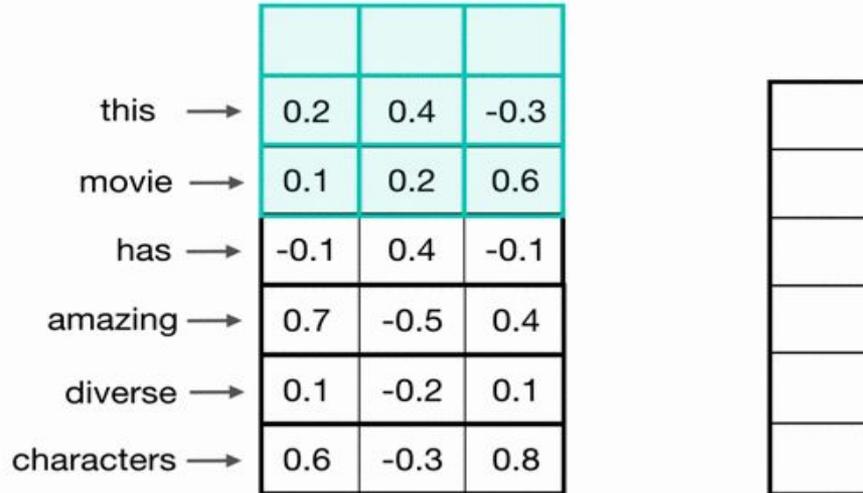


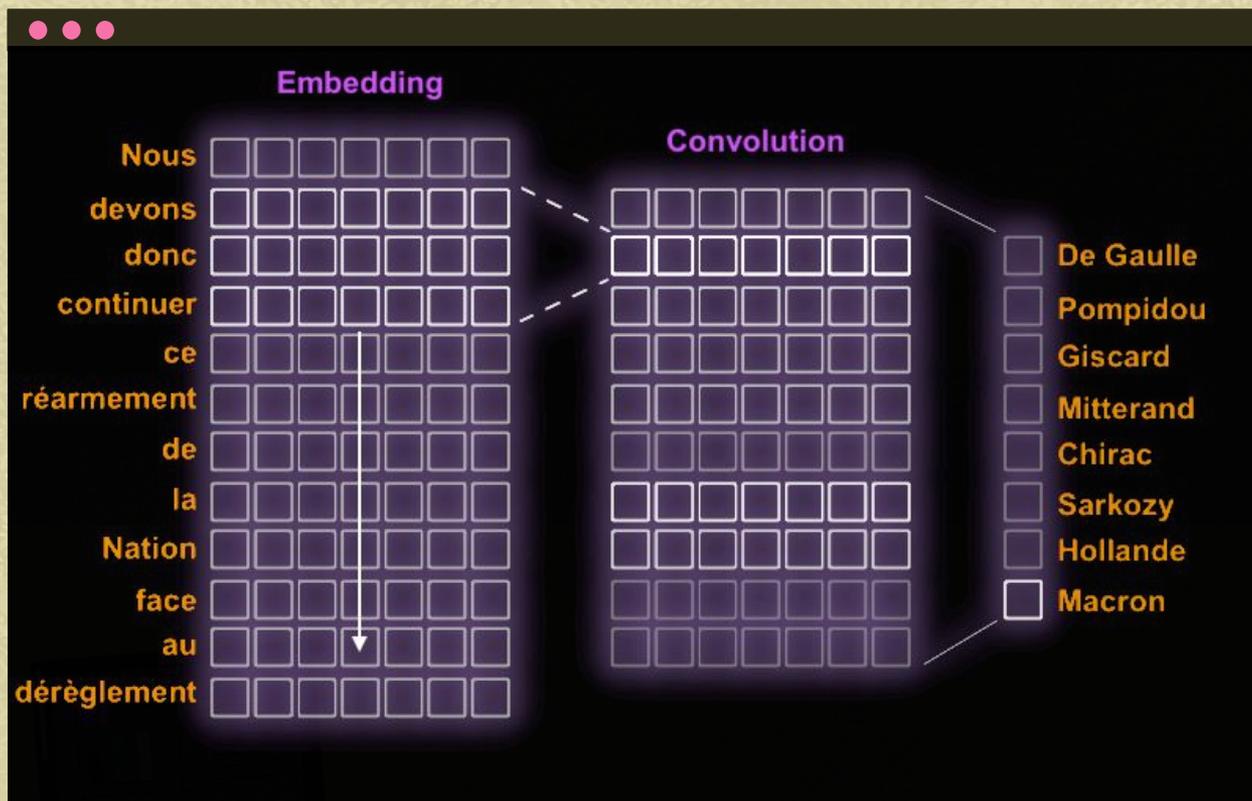
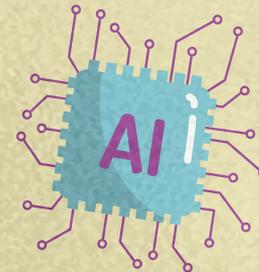
Démo Word2Vec

Entraînement d'une vectorisation de mots

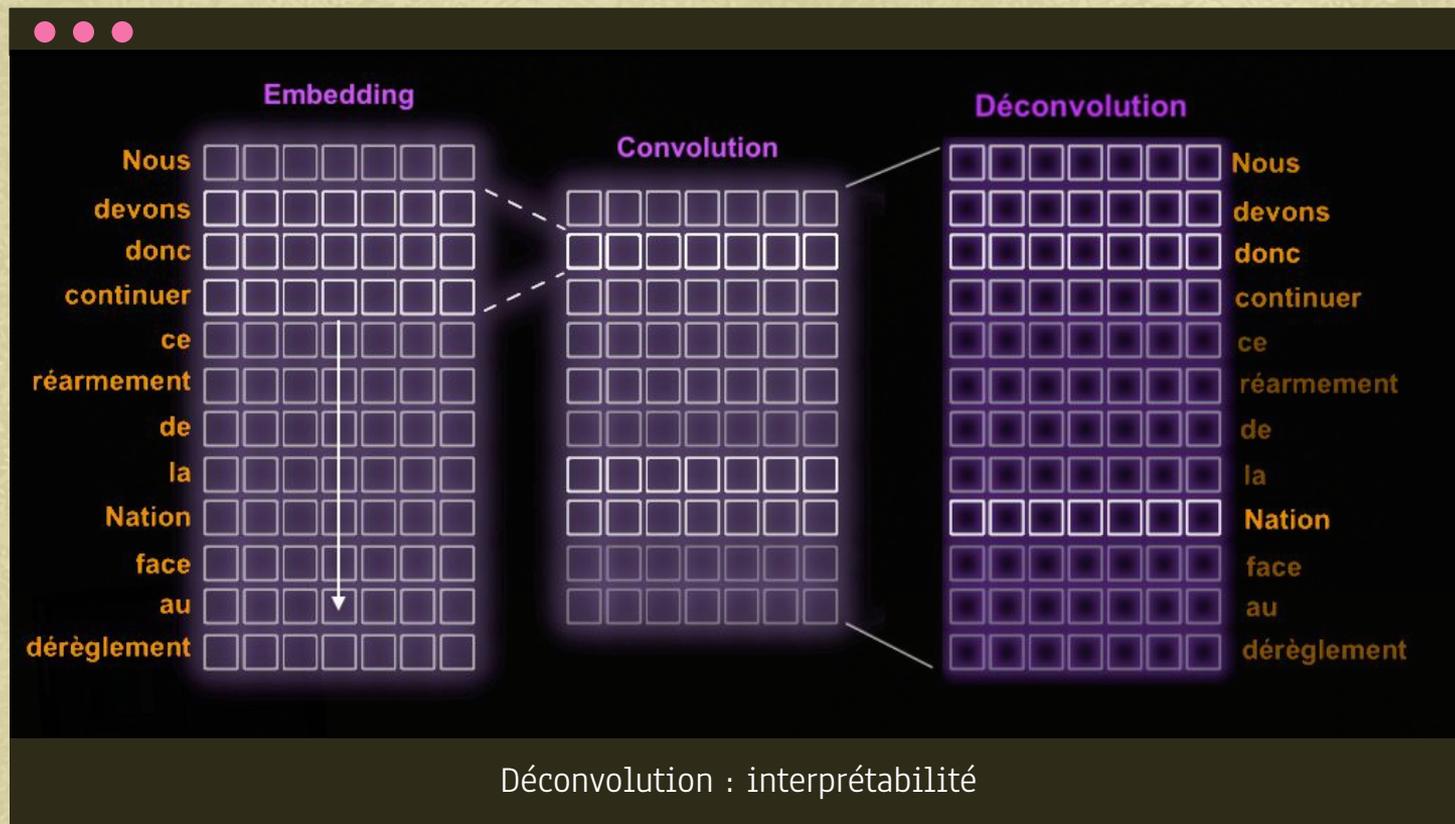
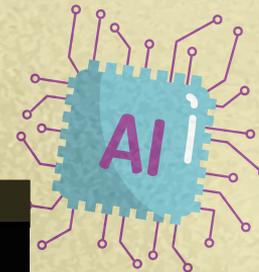
- <https://ronxin.github.io/wevi/>

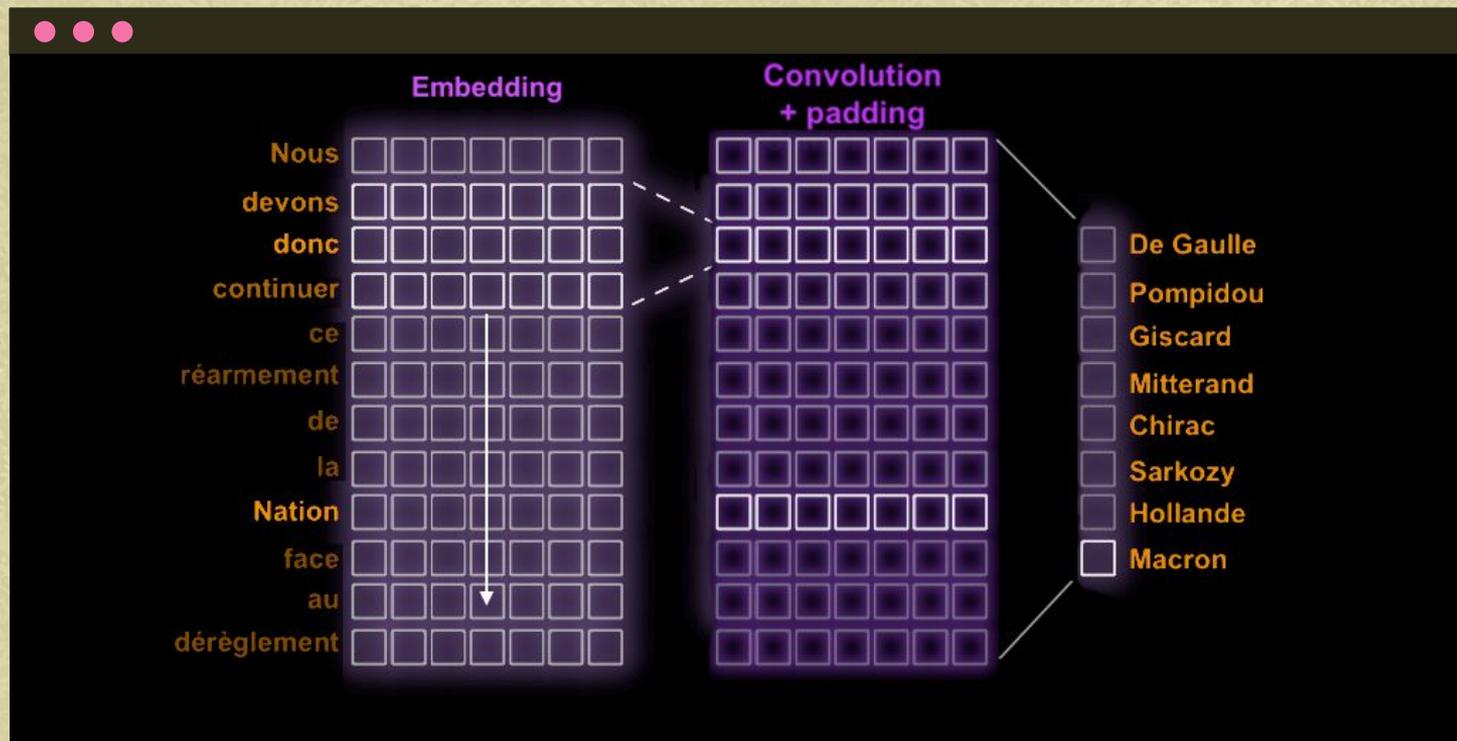
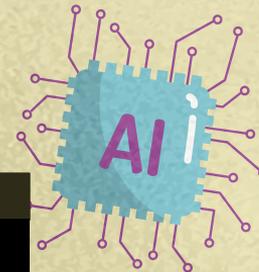
Techniquement : La Convolution



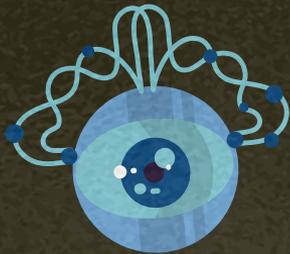


Convolution : exemple de classification



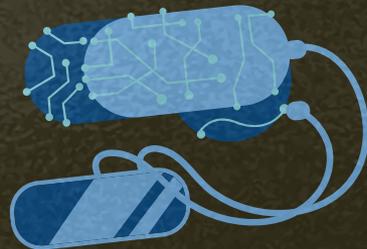


Réseaux interprétables sans déconvolution



Self-Attention

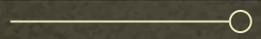
Capture des associations
locales et **distantes** entre les mots





Associations ou paradigmes ?





a) Associations locales b) Associations distantes



L'attention sur le texte. Prochain mot ?

L'école est un lieu animé où chaque matin, les élèves entrent dans leur classe, impatients de commencer la journée. Ils s'assoient, sortent leurs crayons, et regardent le tableau blanc où le professeur inscrit les leçons du jour. Les élèves, attentifs, posent des questions et prennent des notes, transformant le tableau en un espace de découverte. Chaque crayon tracé marque un pas de plus dans leur apprentissage, guidé par la passion et la sagesse de leur



Axe
paradigmatique



Axe
syntagmatique



Prédiction

professeur,
maître, maîtresse



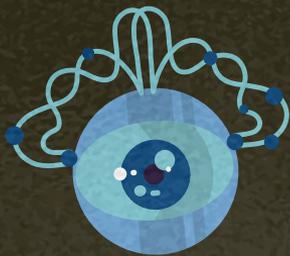
Exemples d'usages



Tâche 1 : Quel est le mot le plus probable pour compléter un texte ?

Tâche 2 : Quelle est la classe la plus probable d'un texte ?

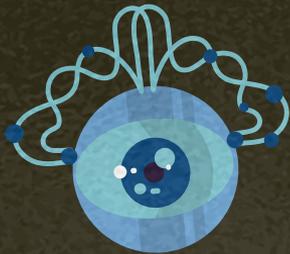




Comment générer un texte à partir de self-attention ?

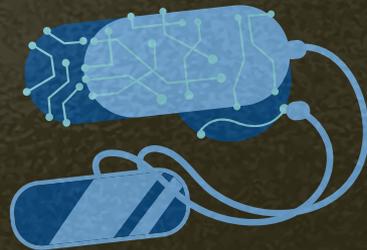
Par itération : Chaque nouveau mot est inséré dans un nouveau "prompt" : l'amorce en entrée du réseau.



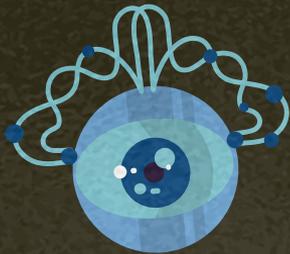


Exemple

Représentation de la construction d'un texte de proche en proche à partir de l'amorce : "Le chat mange"



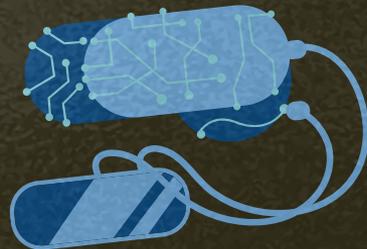


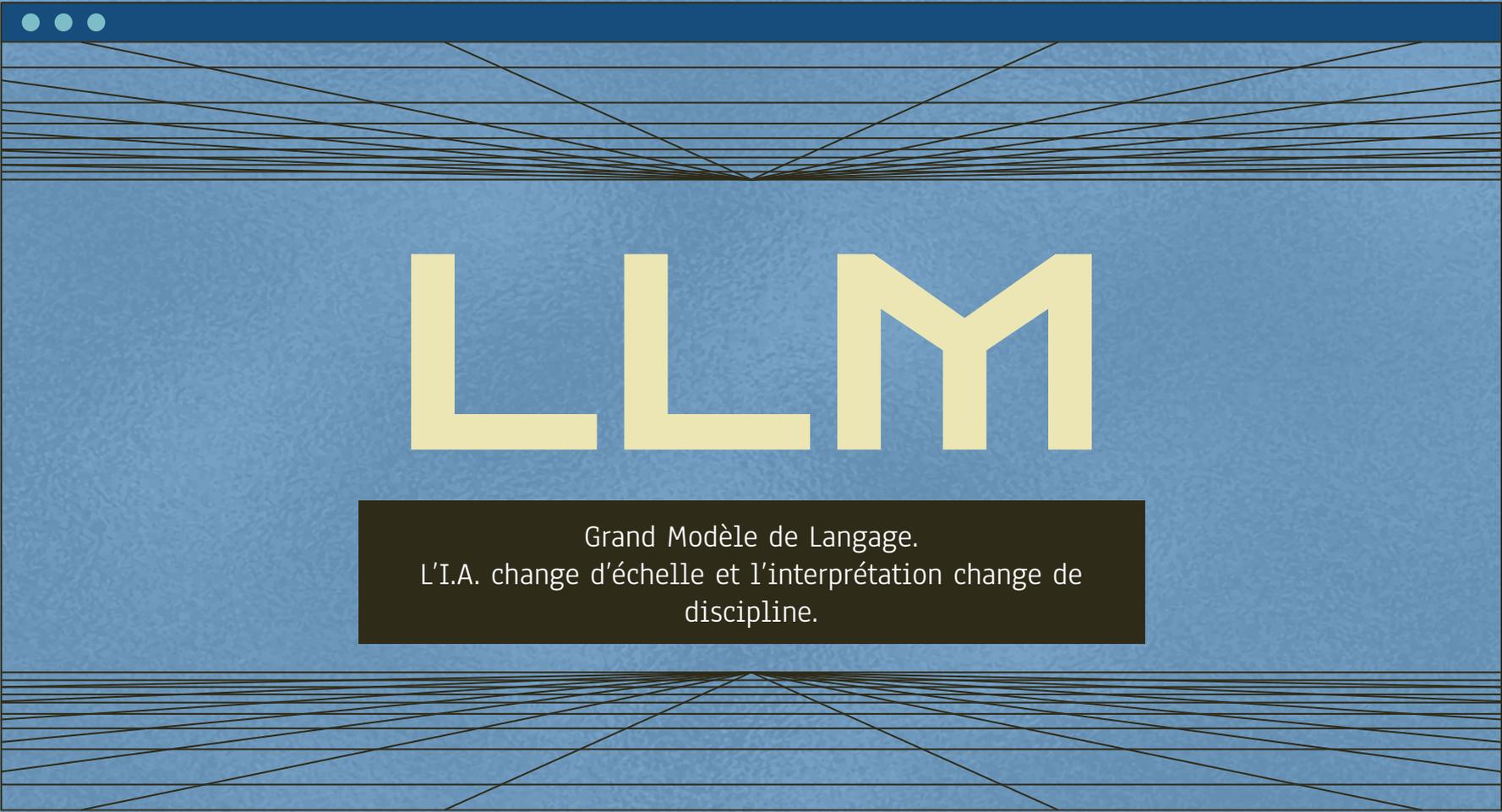


Résultat

“Le chat mange la souris qu’il a attrapée” est une sortie machine qui **représente un texte probable.**

En aucun cas un fait (scientifique) ou une vérité.





LLM

Grand Modèle de Langage.
L'I.A. change d'échelle et l'interprétation change de
discipline.



Chat Grand Public Trompe-l'oeil (GPT)



Dans le crépuscule doré d'une Rome éternelle, le pape François, vêtu d'une doudoune simple mais élégante qui contrastait avec la solennité des siècles passés, avançait parmi ses fidèles avec cette bienveillance chaleureuse et cette humilité touchante qui, malgré la modernité de son habit, irradiaient cette même aura de miséricorde et de réconfort que l'on attend d'un berger des âmes.

(Proust, 2024) par ChatGPT

Chat Generative Pre-trained Transformer

Modèles très profonds

Transformers : **multihead-Attention**

Optimisations

Prompt bien choisi

Données

Des volumes vertigineux

Renforcement

Corrections permanentes

Resultats : **des traitements automatiques du langage bluffants...**

LLM

Traitement automatique +1

Interprétabilité 0

Explicabilité -1



MERCI

Questions?

Laurent.Vanni@univ-cotedazur.fr

Rendez-vous sur : <https://hyperbase.unice.fr>



UMR7320: Bases, Corpus, Langage

Intelligence

Artificielle & Langage

2

Laurent.Vanni@univ-cotedazur.fr



CNRS - Université Côte d'Azur
UMR 7320 : Bases, Corpus, Langage

Deep learning Pour qui ? Pourquoi ?

Sciences

Des nouveaux **observables**
sur le monde.



Marketing

Des modèles **bluffants.**
Pas des I.A. fortes.



Applications

Des outils **puissants.**
Pas des oracles.



Quelles disciplines ? Quelles tâches ?

Sciences du langage

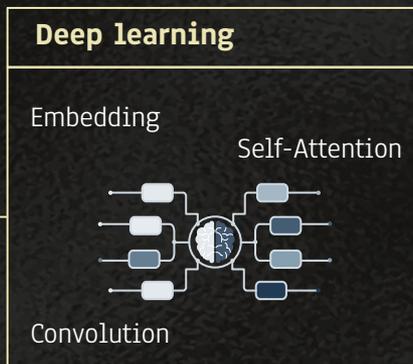
**Analyse de Données Textuelles
(ADT)**



I.A. et ADT : Trois fondamentaux

Corpus

Hypothèses contrastives
Matérialité textuelle
Sciences de la culture



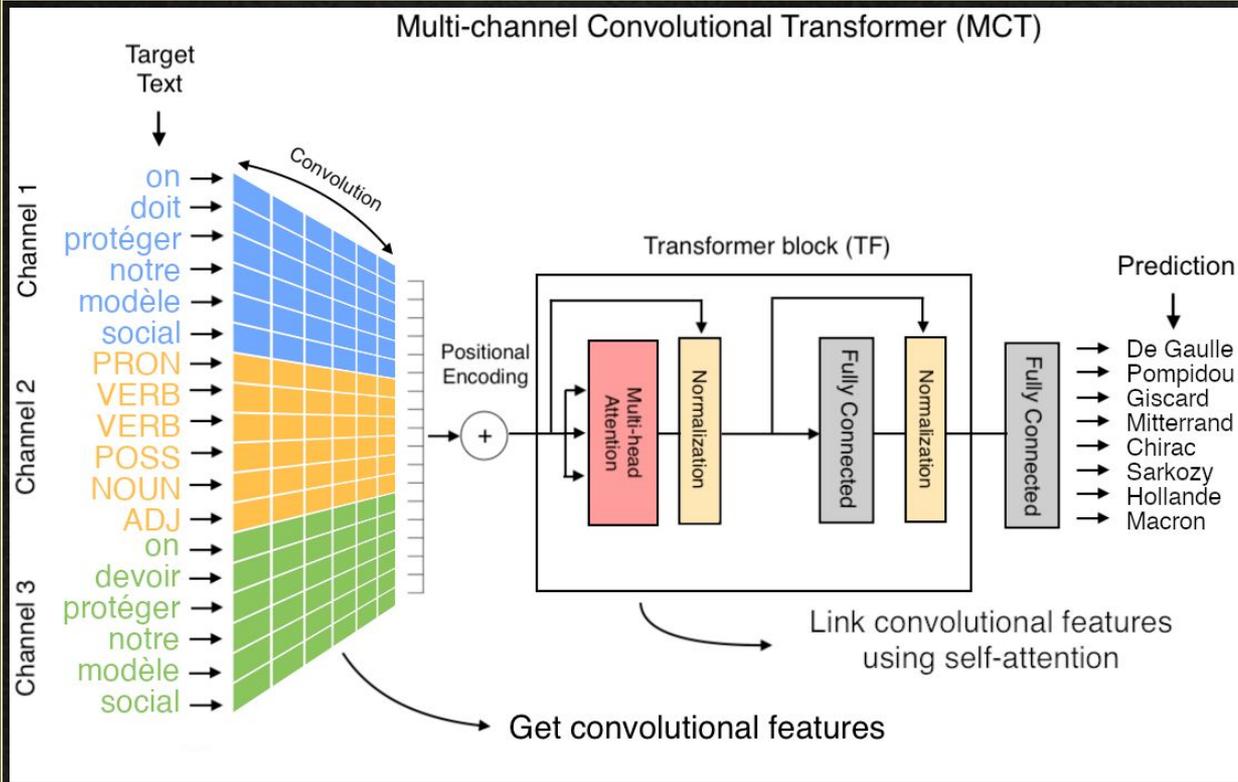
Prédiction

Sortie logicielle attribution
d'un texte à une classe

Description

Interprétation détection de
nouveaux observables ?

Quel modèle ? Pour prédire/décrire

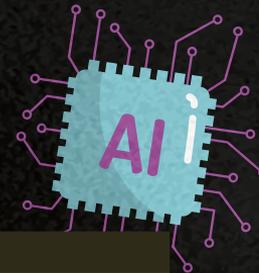


Convolution:
Détection de saillances locales

Axe: syntagmatique

Transformer (self-attention):
Détection d'associations à distances

Axe: paradigmatic



Prédiction : macron 82%

voir aussi	taux d'attribution
chirac	9%
sarkozy	7%
degaulle	1%
giscard	1%
hollande	0%
mitterrand	0.0%
pompidou	0.0%

Passages-clés

[...] **marché européen de le** électricité : toutes ces **grandes transformation ont été décidées** , **VERB** , et **ont commencé à VERB** en **vigueur cette NOUN** . **CCONJ nous avoir** continué cette année de **VERB** des emplois , tout en **VERB** encore plus rapidement nos **émissions de gaz à** effet [...]

voir dans le texte ◀ précédent | 2 | suivant ▶ plus d'infos

Texte intégral

Françaises , **NOUN** , de métropole , de nos **NOUN CCONJ ADP** l' étranger ,

En vous **présentant mes vœux** , je **veux commencer** par exprimer notre **gratitude** à tous pour **nos compatriotes** qui , ce soir encore , **protègent** , **soignent** , aident . **VERB** la continuité de la vie de **la Nation** . **NOUN ADP** eux , nous pouvons **VERB** ces **moments d'** une année à l' **ADJ** en famille et **avec nos proches** . **CCONJ** en **ces instants** , je **veux ADV** dire particulièrement mon affection **ADP** ceux qui **AUX NOUN ou resteront seuls** .

Seuils marqueurs

Forme

Part-of-speech

Lemme

Attention

Motifs profonds

Motifs	Score
CCONJ [...] du marché européen ADP le [...] ces grandes transformation ont [...] VERB	5.43
[...] ont commencé [...]	
vigueur cette	

Hyperdeep démo : attribution du discours des voeux 2024 d'E. Macron

Vanni L., Mahmoudi H., Longrée D. and Mayaffre D. (2024-in press). **Multi-channel Convolutional Transformer** and intertextuality: a Latin case study. In Giordano G. and Misuraca M. (Eds.) *New Frontiers in Textual Data Analysis*, Springer.



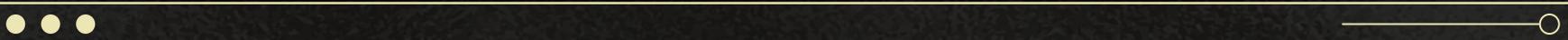
Conclusions Les plus-values de l'I.A.

Nouvelles sorties machines

- + Amélioration des performances en TAL
- Interprétations des sorties machines difficiles
- Généralisations des tâches difficiles

Nouvelles interrogations sur les textes

- + Qu'est-ce qu'un corpus ?
- ++ Que nous apprend le matérialisme textuel ?
- +++ Quels sont les nouveaux observables linguistiques ?



MERCI!

Questions?

Laurent.Vanni@univ-cotedazur.fr
Rendez-vous sur : <https://hyperbase.unice.fr>

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

UMR7320: Bases, Corpus, Langage