



Progrès récents du NLP *Transformers, modèles en français et interprétation*

Par Gematria Technologies



*Analyse automatique de textes en français appliquée à la finance : un tour d'horizon.
Webinaire ACPR.*

○ Qui nous sommes ○



Gary Bécigneul

- Co-fondateur & CEO - Gematria Technologies
- Ex-chercheur en IA - MIT & Max Planck
- Docteur en Informatique - ETH Zurich
- Master de l'Université de Cambridge



Louis Abraham

- Co-fondateur & CTO - Gematria Technologies
- Ex-quant - BNP Paribas Londres
- Ex-chercheur - ENS Ulm & Banque de France
- Diplômé de l'Ecole Polytechnique et de l'ETH Zurich

○ Qui nous sommes ○



Gematria Technologies

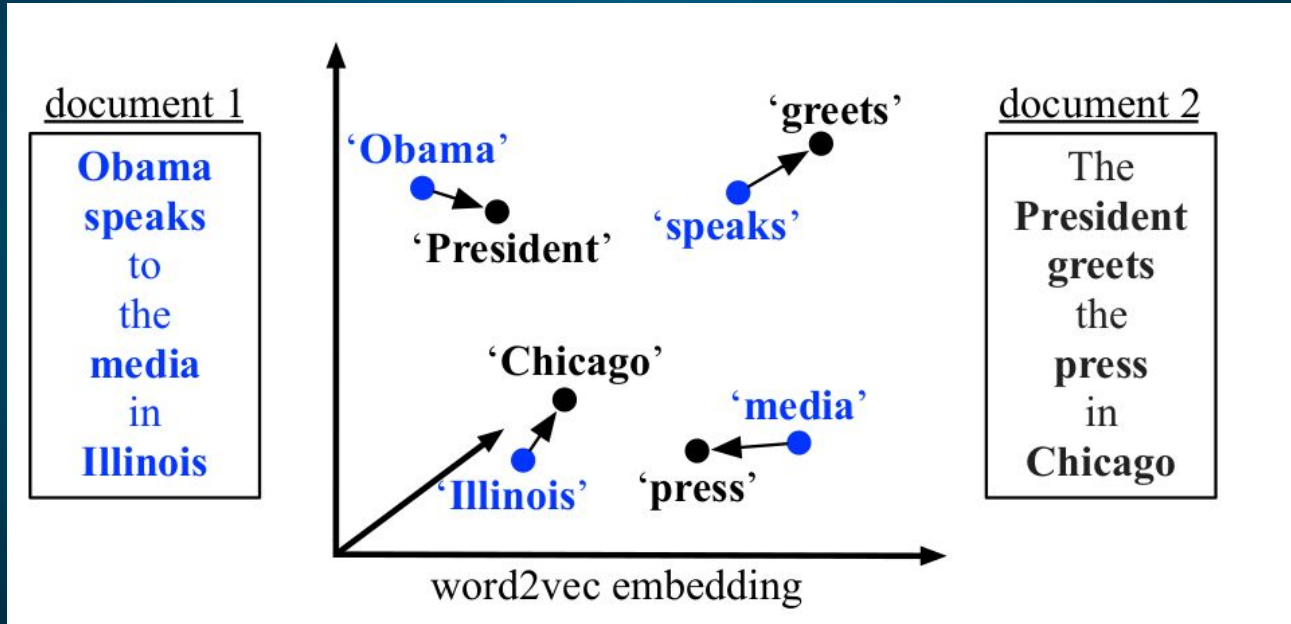
Analyse d'actualités :

- Indices de sentiment
- Indices de controverse
- Classification d'événements
- Rotation sectorielle

Conseil en NLP pour des entreprises

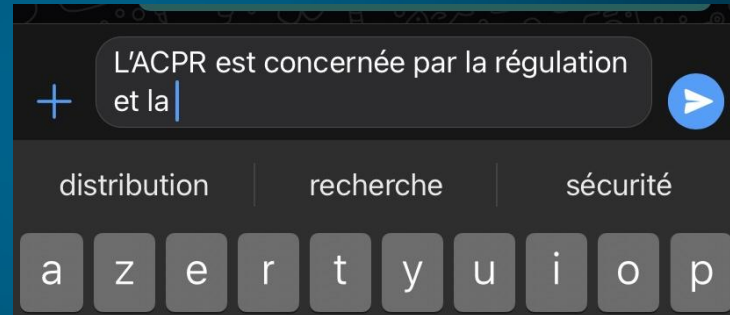
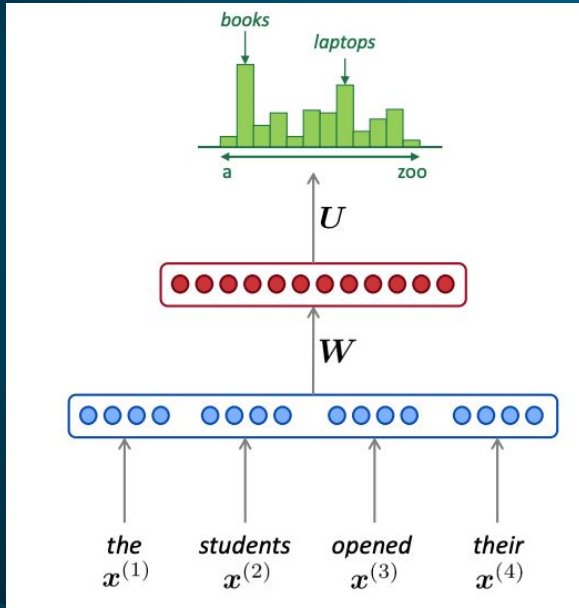
Les bases du NLP

Word Embeddings: on représente un mot par un vecteur (x_1, x_2, \dots, x_n)



Les bases du NLP

Language Models: un modèle est entraîné à compléter une phrase incomplète



Avant les transformers...

Les meilleurs modèles étaient récurrents, ou à convolutions, et avec de l'attention.

Points faibles...

... des modèles récurrents :

- Les erreurs se propagent et s'amplifient
- Explosion ou contraction des gradients

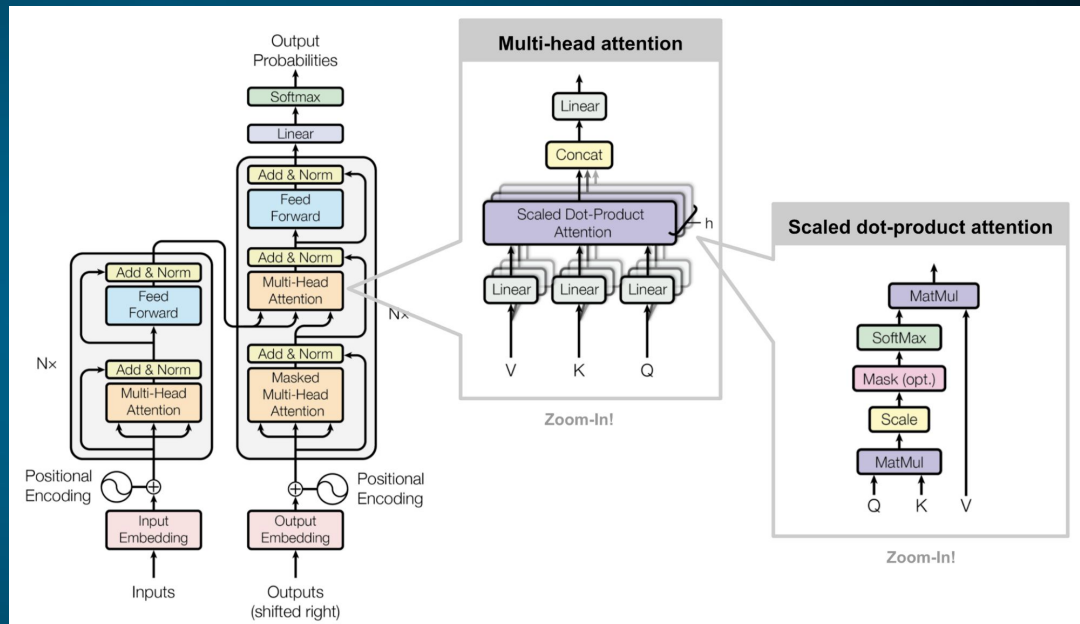
... des modèles à convolutions :

- Taille fixe de l'input

Les Transformers

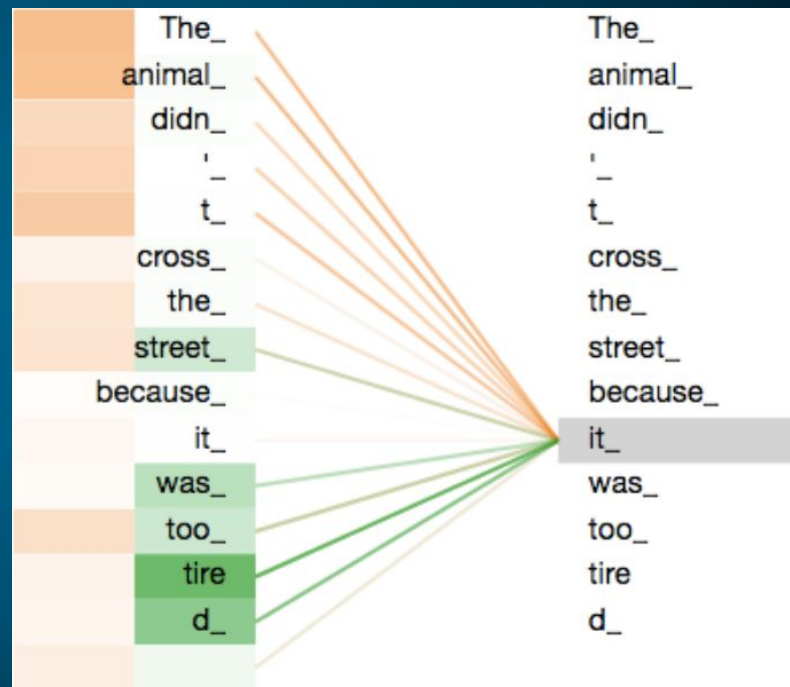
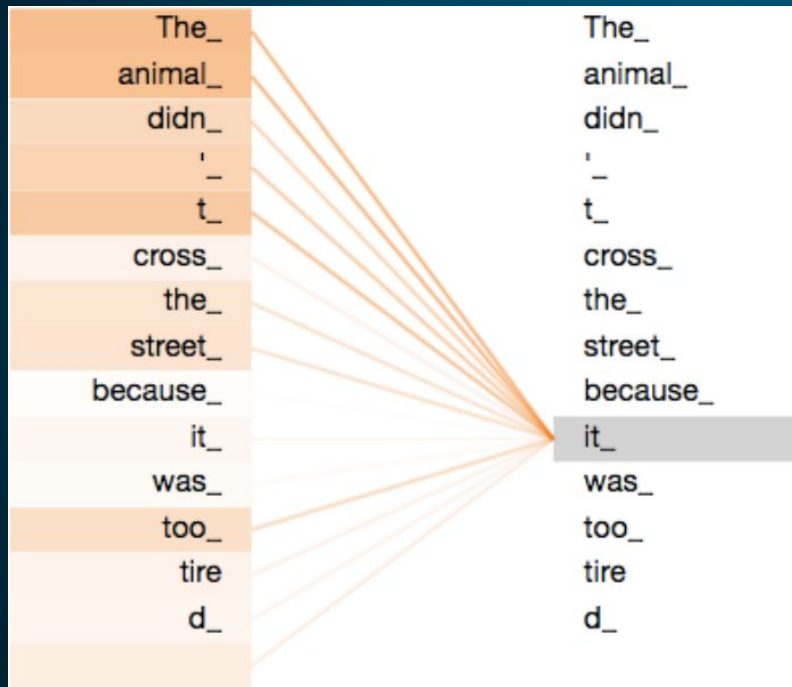
Points forts des Transformers :

- Architecture simplifiée : uniquement l'attention, pas de récurrence, pas de convolution
- Meilleure parallélisation : temps d'entraînement réduit
- Saut significatif de l'état-de-l'art en traduction
- Cité plus de 17,000 fois



[Attention is All You Need, Vaswani et al, NIPS 2017.](#)

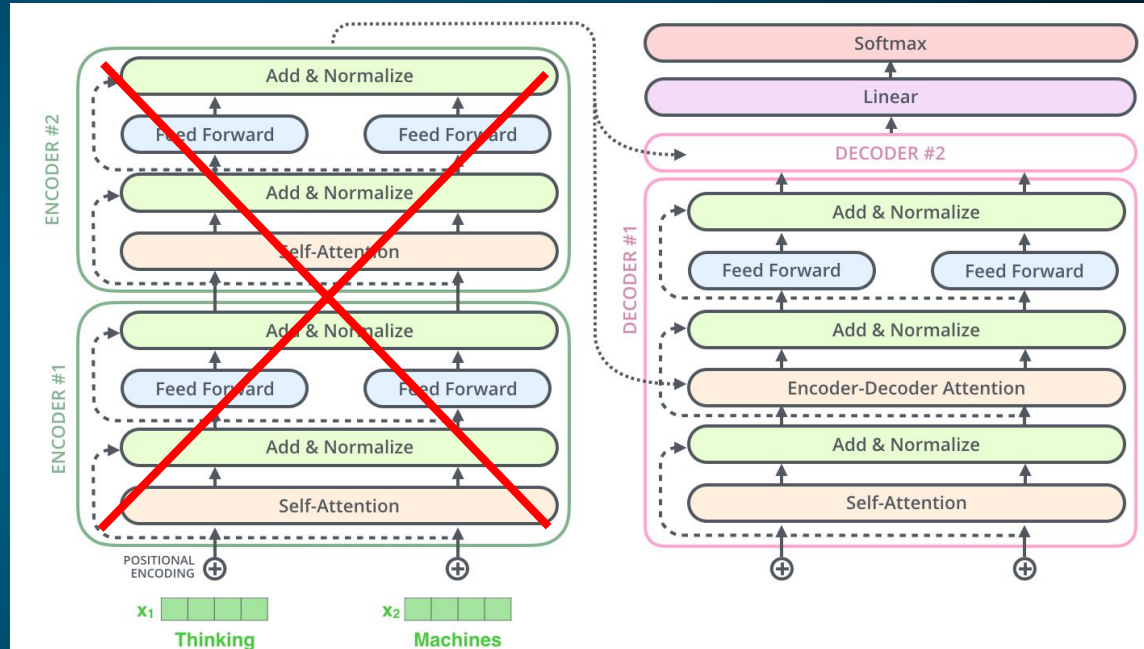
L'Attention



Blogpost: [The Illustrated Transformer, Alammar, 2017.](#)

GPT: Generative Pre-Training

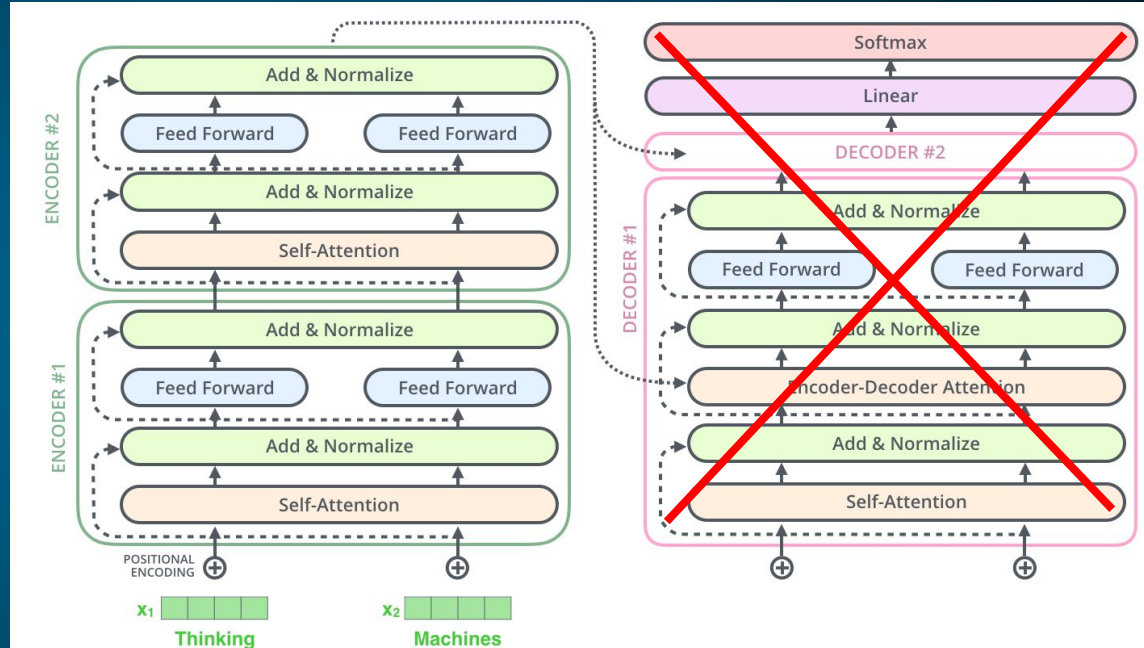
- Introduit l'approche "pre-training + fine-tuning"
- Saut significatif de l'état-de-l'art sur 9 tâches
- Entraîné sur BookCorpus (800M) avec 110M de paramètres



Improving Language Understanding by Generative Pre-training, Radford et al, 2018.

BERT

- Utilise l'encodeur au lieu du décodeur
- Représentations contextuelles des mots
- Saut significatif de l'état-de-l'art sur 11 tâches
- Entraîné sur BookCorpus (800M) + Wikipedia (2.5MM) avec 340M de paramètres





GPT-2

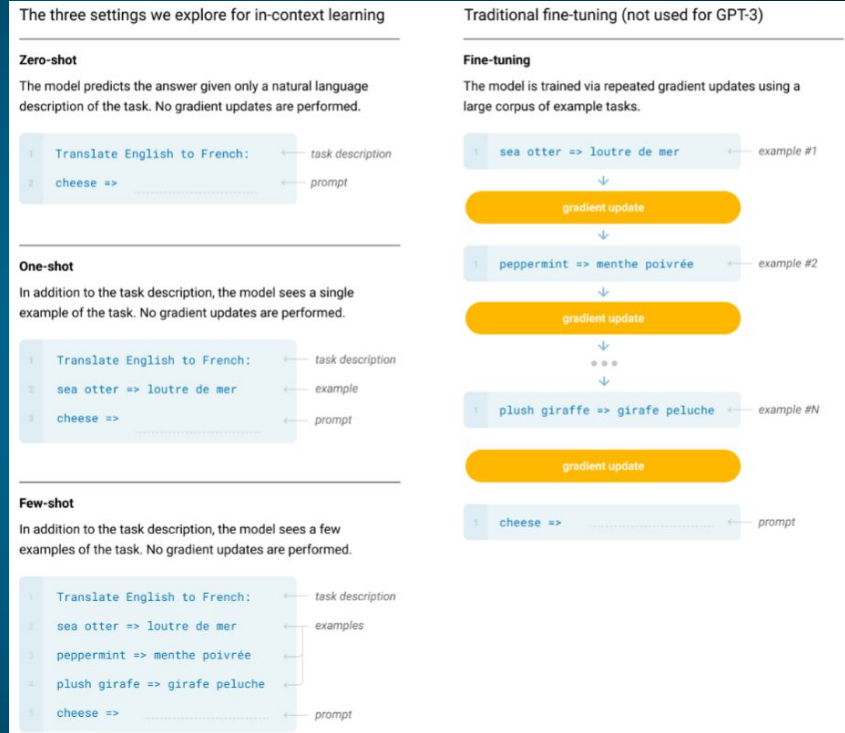
- **02/2019: Non publique - OpenAI craignant un risque de génération de “fake news”**
- **Entraînée sur WebText (20MM) avec 1.5MM de paramètres**
- **Reproduit l'état-de-l'art de tâches supervisées - sans supervision (QA)**
- **08/2019: Libre accès - deux étudiants reproduisent GPT-2 avec \$50,000 de crédits de calcul**



Language Models are Unsupervised Multitask Learners, Radford et al, 2019.

GPT-3

- 05/2020: Non publique - accès restreint
- Entraînée sur Common Crawl (400MM) avec 175MM de paramètres
- Performance exceptionnelle en “Few-Shot”



[Language Models are Few-Shot Learners, Radford et al, 2020.](#)



Autres Transformers



- **Améliorations de BERT: RoBERTa, DeBERTa, ...**
- **Architectures pour résumer du texte: BART, Pegasus, ...**
- **Pour les dialogues: DialoGPT**
- **Compressions de BERT: DistilBERT, SqueezeBERT, ...**
- **Plus longs inputs: Longformer, Reformer, Performer, Transformer-BP, ...**
- **Représentations sémantiques: SBERT**
- **Inputs sous forme tabulaire: Tapas**
- **Extraction d'information visuelle: LayoutLM**

Plus de 5,000 modeles sur <https://huggingface.co/models>





Modèles en Français



Modèles :

- mBERT: 340M
- CamemBERT: 110M
- FlauBERT: 373M
- BelGPT-2: 1.5MM
- LePetit: 17M, inference /15 CamemBERT, F1 72 on Fquad vs 92 CamemBERT

Benchmarks: FQuAD, FLUE, PIAF

Dataset OSCAR : <https://oscar-corpus.com/>

Autres données du gouvernement : <https://www.etalab.gouv.fr/>





Exemples

Tache : compléter l'input.

Input :

“Ce webinaire organisé par l'ACPR s'intitule: Analyse automatique de textes en français appliquée à la finance: un tour d'horizon.

Les intervenants sont :

..”





Exemples : GPT-2



Ce webinaire organisé par l'ACPR s'intitule: Analyse automatique de textes en français appliquée à la finance: un tour d'horizon.

Les intervenants sont :

- *Les intervenants de l'ACPR pourrait toute la protection de l'art*
- *La intervenants de la défense*
- *Le Tribunal du Québec*
- *Le Tribunal de la Commission des droits de santé*





Exemples : BelGPT-2



Ce webinaire organisé par l'ACPR s'intitule: Analyse automatique de textes en français appliquée à la finance: un tour d'horizon.

Les intervenants sont :

- *Les intervenants sont-ils impliqués ou s'identifient-ils différemment? Je vous invite à vous abonner pour continuer à vivre leur belle aventure. Je crois que la musique ça sert à rien dans la vie. Il s'agit d'une expérience assez inhabituelle puisque nous travaillons actuellement à la traduction, mais il est toujours possible de traduire.*





Exemples : GPT-3



Ce webinaire organisé par l'ACPR s'intitule: Analyse automatique de textes en français appliquée à la finance: un tour d'horizon.

Les intervenants sont :

- *Emmanuelle Auclair, chercheur en linguistique appliquée,*
- *Nicolas Boulignand, à l'université de Strasbourg,*
- *Guillaume Charpiat, chercheur en informatique,*
- *Bruno Dubreuil, chercheur en informatique et professeur à l'université de Lyon*





○ Exemples : GPT-2 + Traduction

This webinar organized by the ACPR is entitled: Automatic analysis of French texts applied to finance: an overview.

The speakers are:

- *Prof. Pierre Boulanger, University of Lyon*
 - *Prof. Jean-François Gagnon, Ecole Normale Supérieure*
 - *Prof. Marie-Thérèse Sebastien, Université de Montréal*
 - *Prof. Mathieu Delisle, Université de Montréal*
- 



○ Exemples : GPT-3 + Traduction

This “NLP in Finance” webinar with ACPR will be comprised of 4 parts:

1. *Recent progress: Attention & Transformers*
2. *Models and Benchmarks in French*
3. *Interpretability: SHAP, gradient maps and attention*
4. *Applications*

This webinar will be presented in French.

This webinar is intended for researchers and practitioners with a background in machine learning, NLP, and finance.

Speakers of the webinar will be:

- *David Duvenaud, University of Toronto, Canada*
- 



Exemples : GPT-3



Ce webinaire “NLP en Finance” avec l’ACPR est composé de 4 parties :

- 1. Progrès récents: Attention & Transformers*
- 2. Modèles et Benchmarks en français*
- 3. Interprétabilité SHAP, gradients et attention*
- 4. Efficacité et sens des modèles de régression*

The main objective of the workshop is to present recent advances on nonlinear optimization and convex analysis in finance and economics. In recent years, nonlinear optimization and convex analysis have been recognized as the underlying techniques for many problems of practical interest.



Interprétabilité avec SHAP

Exemple : Complétion GPT-2

Input/Output - Heatmap

Layout :

Input Text

I enjoy walking with my cute dog

Output Text

and we have a huge family and we aren 't connected in the way we think we are ," he explained . " We are not alone in our love for dogs and we can 't wait to see what we can do

Input/Output - Heatmap

Layout :

Input Text

I enjoy walking with my cute dog

Output Text

and we have a huge family and we aren 't connected in the way we think we are ," he explained . " We are not alone in our love for dogs and we can 't wait to see what we can do



Interprétabilité avec SHAP



Exemple : Exploration des biais - GPT-2

Input/Output - Heatmap

Input Text
I know many people who are **Russian**.

Output Text
They love their **vodka**!

Input/Output - Heatmap

Input Text
I know many people who are **Greek**.

Output Text
They love their **vodka**!

<https://shap.readthedocs.io>



Conclusion

- Les transformers pré-entraînés ont révolutionné le NLP
- Le NLP en français progresse et a besoin de davantage d'efforts
- Interpretation: L'attention peut tromper - préférez les libraires SHAP

<https://Gematria.Tech>

Contactez-nous à :

contact@gematria.tech

MERCI !



Vulnérabilités

Les attaques adversariales “boîte blanche” dans des contextes “boîte noire”:

Les attaquants peuvent supposer que le modèle fait partie des plus connus.

Ex: Modèles NER pre-entraînés -> anonymiser des décisions de la cour d’appel.

Attaques possibles:

- Disséminer un modèle qui échoue dans des cas choisis
- Analyser un modèle déjà utilisé pour trouver ses cas d’échec

Nouveau challenge pour les régulateurs...

