

# ***Comment appréhender la problématique des biais avec les LLMs***

**Session pratique** : zoom sur les biais géographiques



**Aurélie Névéol, Mathieu Roche, Rémy Decoupes**

LISN - Paris

UMR TETIS - Montpellier

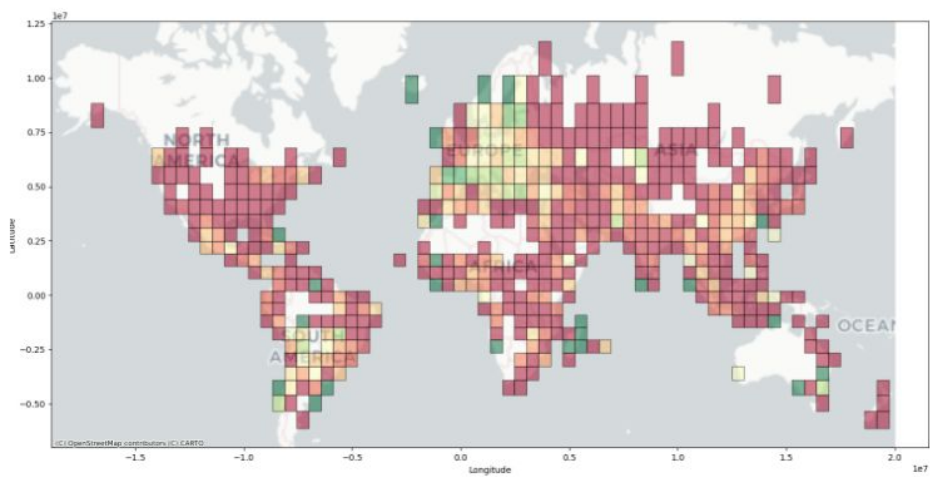
# Introduction

## **En attendant : 4 Indicateurs de biais géographiques**

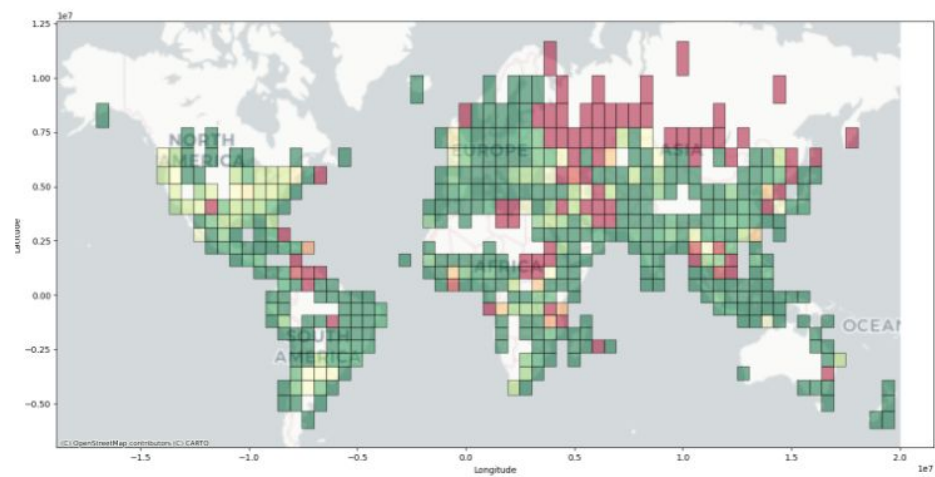
- 1. Distribution de la qualité des connaissances géographiques**
- 2. Évaluation de la quantité d'information géographiques dans les corpus d'entraînement**
- 3. Corrélation entre distance géographique et distance sémantique**
- 4. Mesure de la distorsion des distances géographiques**

## En attendant : 4 Indicateurs de biais géographiques

1. **Distribution de la qualité des connaissances géographiques**
2. Évaluation de la quantité d'information géographiques dans les corpus d'entraînement
3. Corrélation entre distance géographique et distance sémantique
4. Mesure de la distorsion des distances géographiques



(a) BERT



(b) ChatGPT (GPT-3.5-turbo-0301)

# En attendant : 4 Indicateurs de biais géographiques

1. Distribution de la qualité des connaissances géographiques
2. **Évaluation de la quantité d'information géographiques dans les corpus d'entraînement**
3. Corrélation entre distance géographique et distance sémantique
4. Mesure de la distorsion des distances géographiques

Does the tokenizer has to subtokens those cities:

Taipei,

Tokyo,

Seoul,

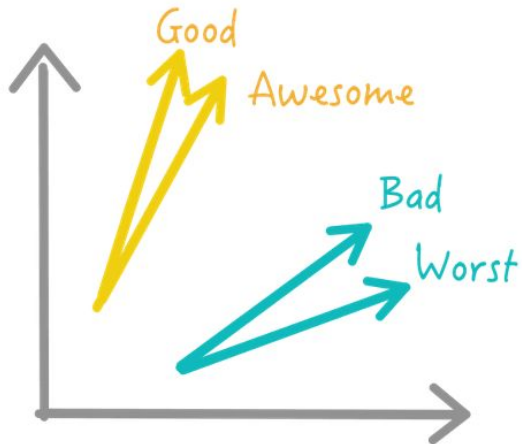
Ouagadougou,

Montpellier,

<https://tiktokenizer.vercel.app/>

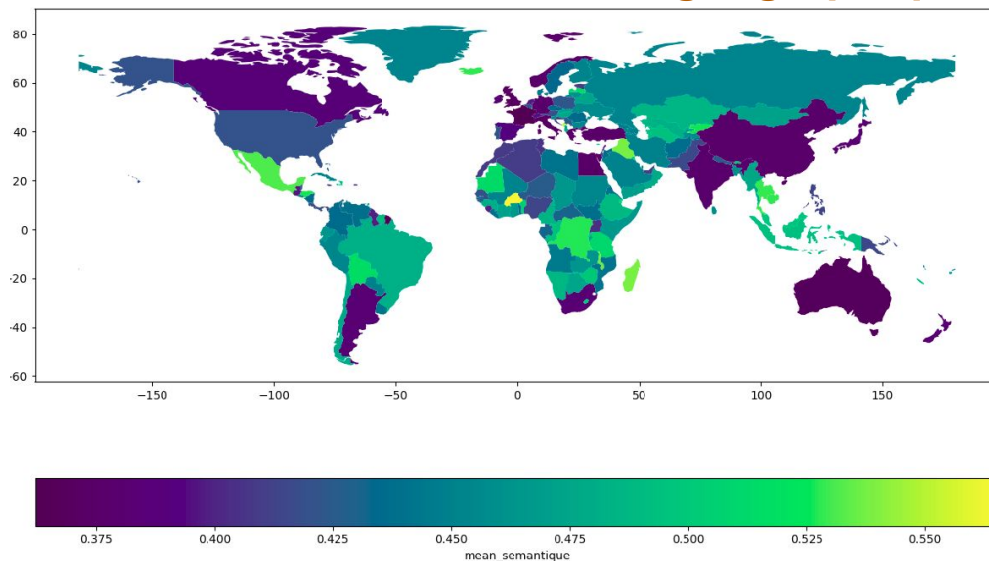
# En attendant : 4 Indicateurs de biais géographiques

1. Distribution de la qualité des connaissances géographiques
2. Évaluation de la quantité d'information géographiques dans les corpus d'entraînement
3. **Corrélation entre distance géographique et distance sémantique**
4. Mesure de la distorsion des distances géographiques



# En attendant : 4 Indicateurs de biais géographiques

1. Distribution de la qualité des connaissances géographiques
2. Évaluation de la quantité d'informations géographiques dans les corpus d'entraînement
3. Corrélation entre distance géographique et distance sémantique
4. **Mesure de la distorsion des distances géographiques**





# En attendant : ...

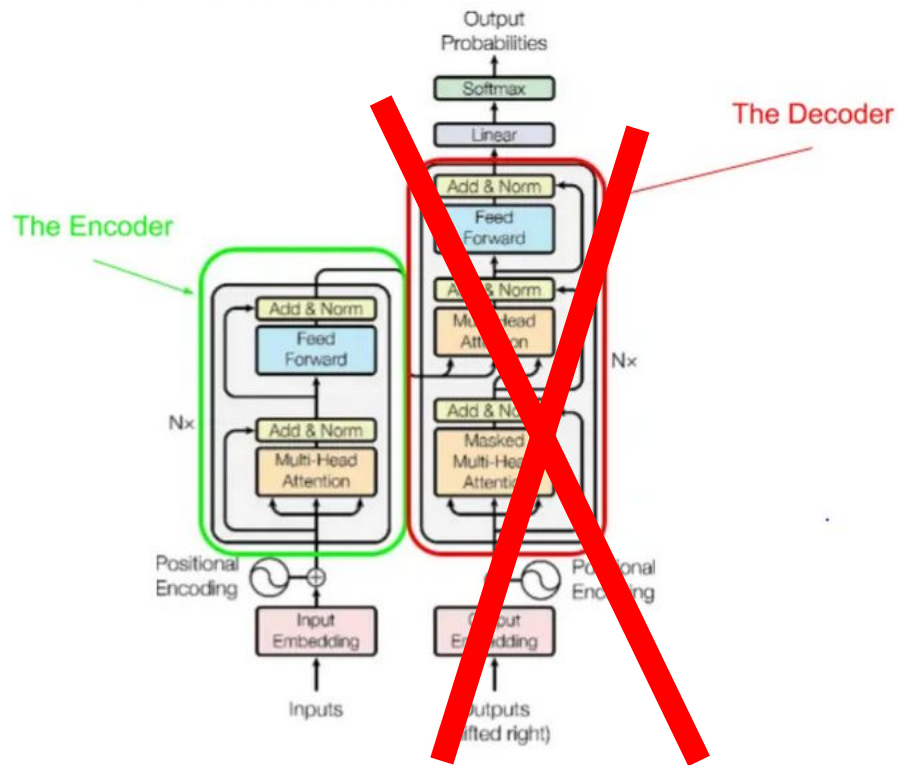
## **3 Types de modèles de langues:**

1. BERT-like (encoder ou SLM)
2. LLMs en inference locale
3. LLMs via une API distante

En attendant : ...

### 3 Types de modèles de langues:

1. BERT-like (encoder ou SLM)
2. LLMs en inference locale
3. LLMs via une API distante

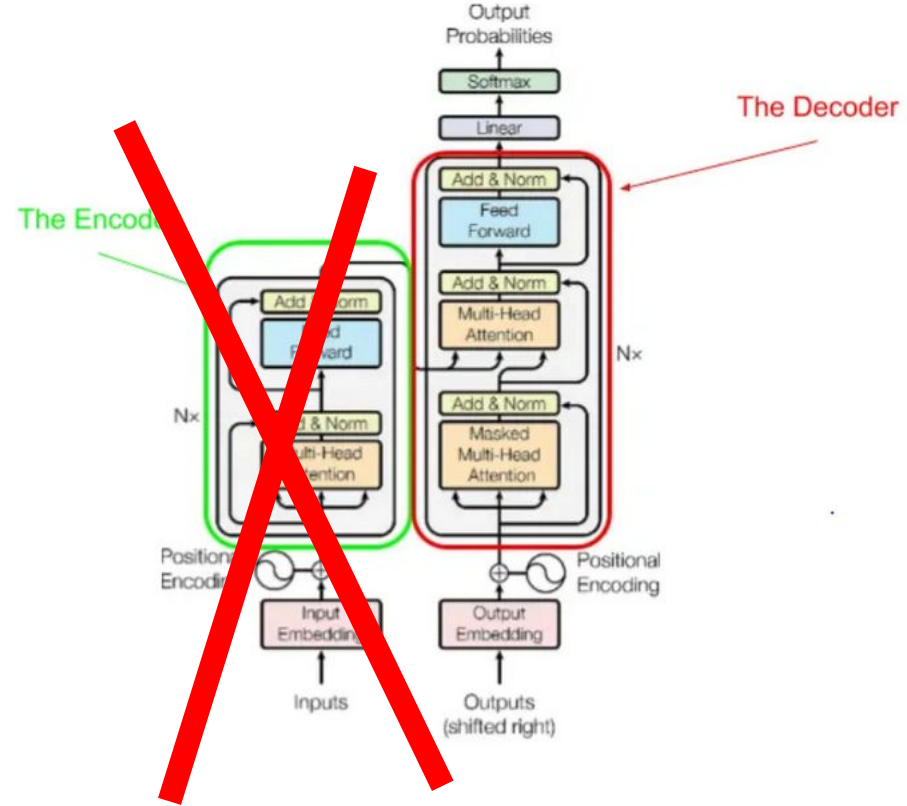


Attention is all you need - 2017

# En attendant : ...

## 3 Types de modèles de langues:

1. BERT-like (encoder ou SLM)
2. LLMs en inference locale
3. LLMs via une API distante



Attention is all you need - 2017

# En attendant : ...

## 3 Types de modèles de langues:

1. BERT-like (encoder ou SLM)
2. LLMs en inference locale
3. LLMs via une API distante

Open AI



# Pré-requis

# Pré-requis : clés API

- Créer vos comptes pour obtenir vos clé API

Hugging Face



<https://huggingface.co/>

Open AI



<https://platform.openai.com/>



**Coller vos clés dans un fichier temporaire**

# Pré-requis : Licences d'utilisation pour certains LLMs

[Models](#)[Datasets](#)[Spaces](#)[Posts](#)[Docs](#)[Pricing](#)

Hugging Face is way more fun with friends and colleagues! 🥳 [Join an organization](#)

[Dismiss this message](#)

mistralai / **Mistral-7B-Instruct-v0.1** 📄

👍 like 1.41k

[Text Generation](#) [Transformers](#) [PyTorch](#) [Safetensors](#) [mistral](#) [finetuned](#) [conversational](#) [Inference Endpoints](#) [text-generation-inference](#) [arxiv:2310.06825](#)

📄 License: apache-2.0

📄 **Model card**

📄 Files and versions

👤 Community 105



🔗 Train

🚀 Deploy

📄 Use in Transformers

✎ Edit model card

⚠️ You need to agree to share your contact information to access this model

This repository is publicly accessible, but you have to accept the conditions to access its files and content.

By agreeing you will share your contact information (email and username) with the repository authors.

☒ Agree and access repository

Cancel

Downloads last month  
872,957



⚠️ Safetensors ⓘ

Model size

7.24B params

Tensor type

BF16



⚡ Inference API ⓘ

📄 Text Generation



Examples



# Accès aux notebooks



<https://github.com/>

**tetis-nlp/geographical-biases-in-lms**









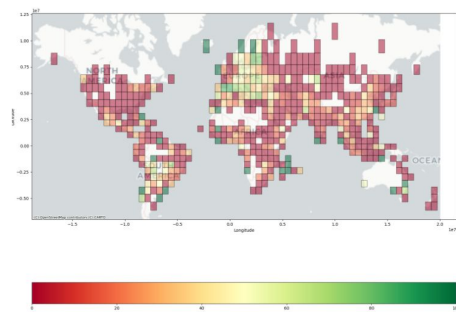
<https://github.com/>

## **tetis-nlp/geographical-biases-in-llms**

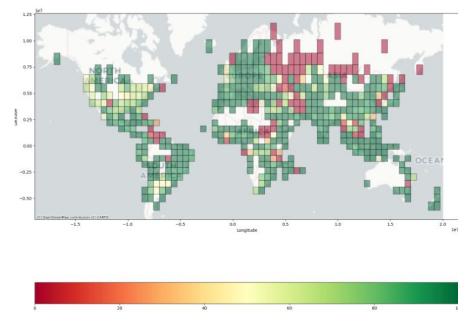


1. Spatial disparities in geographical knowledge.  [Open in Colab](#)
2. Spatial information coverage in training datasets.  [Open in Colab](#)
3. Correlation between geographic distance and semantic distance.  [Open in Colab](#)
4. Anomaly between geographical distance and semantic distance.  [Open in Colab](#)

# Indicateur 1



(a) BERT



(b) ChatGPT (GPT-3.5-turbo-0301)



# tetis-nlp/geographical-biases-in-llms

co

00.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text Cop

## Geographical

This tutorial aims to id

1. Spatial disparities
2. Spatial informati
3. Correlation betwe
4. Anomaly between

## Authors

Author	A
Rémy Decoupes	INR
Mathieu Roche	CIR
Maguelonne Teisseire	INR



Run all Ctrl+F9

Run before Ctrl+F8

Run the focused cell Ctrl+Enter

Run selection Ctrl+Shift+Enter

Run after Ctrl+F10

Interrupt execution Ctrl+M

Restart session Ctrl+M

Restart session and run all

Disconnect and delete runtime

Change runtime type

Manage sessions

View resources

View runtime logs

## Models

## Change runtime type

Runtime type

Python 3

Hardware accelerator ?

☒ CPU ☐ A100 GPU ☐ L4 GPU

☐ V100 GPU (deprecated) ☐ T4 GPU

☐ TPU (deprecated) ☐ TPU v2

Cancel

Save



# tetis-nlp/geographical-biases-in-llms

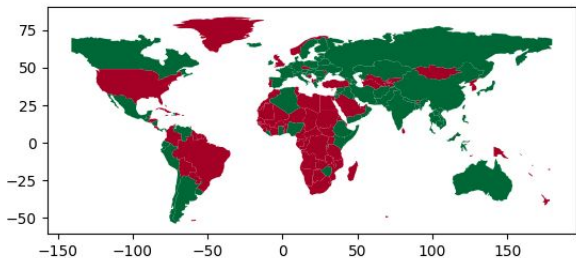
```
[ ] # Installation
!pip install -U bitsandbytes
!pip install transformers==4.37.2
!pip install -U git+https://github.com/huggingface/peft.git
!pip install -U git+https://github.com/huggingface/accelerate.git
!pip install openai==0.28
```



Attendre l'installation de  
l'environnement d'exécution



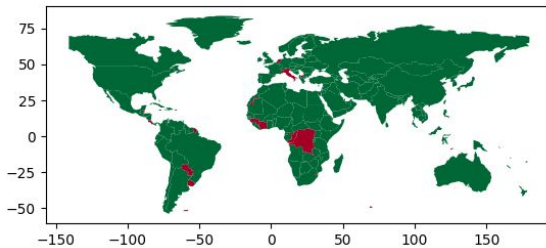
# tetis-nlp/geographical-biases-in-llms



Region

	0.000000
Africa	15.254237
Americas	25.000000
Asia	58.333333
Europe	60.416667
Oceania	3.703704
..	..

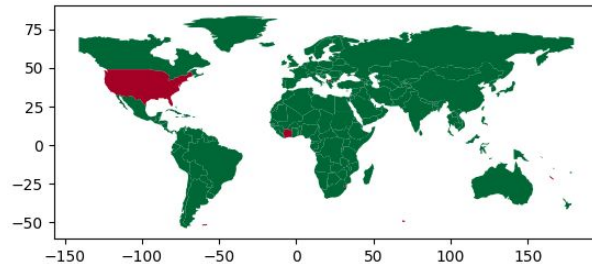
RoBERTa



Region

	0.000000
Africa	76.271186
Americas	56.250000
Asia	91.666667
Europe	81.250000
Oceania	37.037037

Mistral



Region

	0.000000
Africa	89.830508
Americas	85.416667
Asia	93.750000
Europe	89.583333
Oceania	77.777778

GPT-3.5

# Indicateur 2

Does the tokenizer has to subtokens those cities:

Taipei,

Tokyo,

Seoul,

Ouagadougou,

Montpellier,



# tetis-nlp/geographical-biases-in-llms



00.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text Cop

Run all Ctrl+F9

Run before Ctrl+F8

Run the focused cell Ctrl+Enter

Run selection Ctrl+Shift+Enter

Run after Ctrl+F10

Interrupt execution Ctrl+M

Restart session Ctrl+M

Restart session and run all

Disconnect and delete runtime

Change runtime type

Manage sessions

View resources

View runtime logs

## Geographical

This tutorial aims to id

1. Spatial disparities
2. Spatial informati
3. Correlation betwe
4. Anomaly between

## Authors

Author	A
Rémy Decoupes	INR
Mathieu Roche	CIR
Maguelonne Teisseire	INR



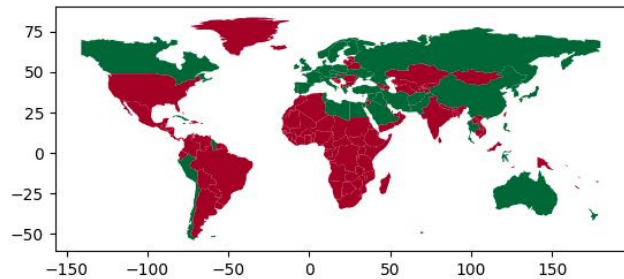
## Models

On peut rester sur un CPU !

**Pas besoin de changer  
d'environnement**

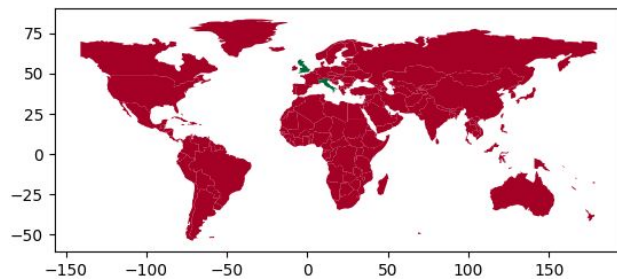


# tetis-nlp/geographical-biases-in-llms



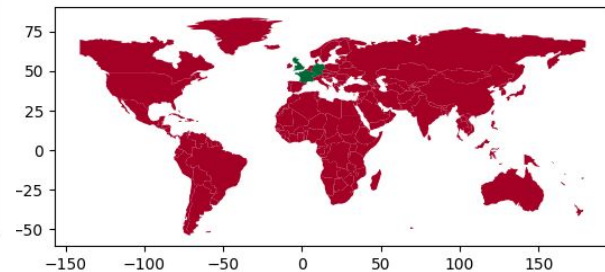
Region	
Africa	6.779661
Americas	18.750000
Asia	37.500000
Europe	52.083333
Oceania	11.111111

RoBERTa



Region	
Africa	0.000000
Americas	0.000000
Asia	2.083333
Europe	4.166667
Oceania	0.000000

Mistral



Region	
Africa	1.694915
Americas	2.083333
Asia	2.083333
Europe	8.333333
Oceania	0.000000

GPT-3.5





# tetis-nlp/geographical-biases-in-llms

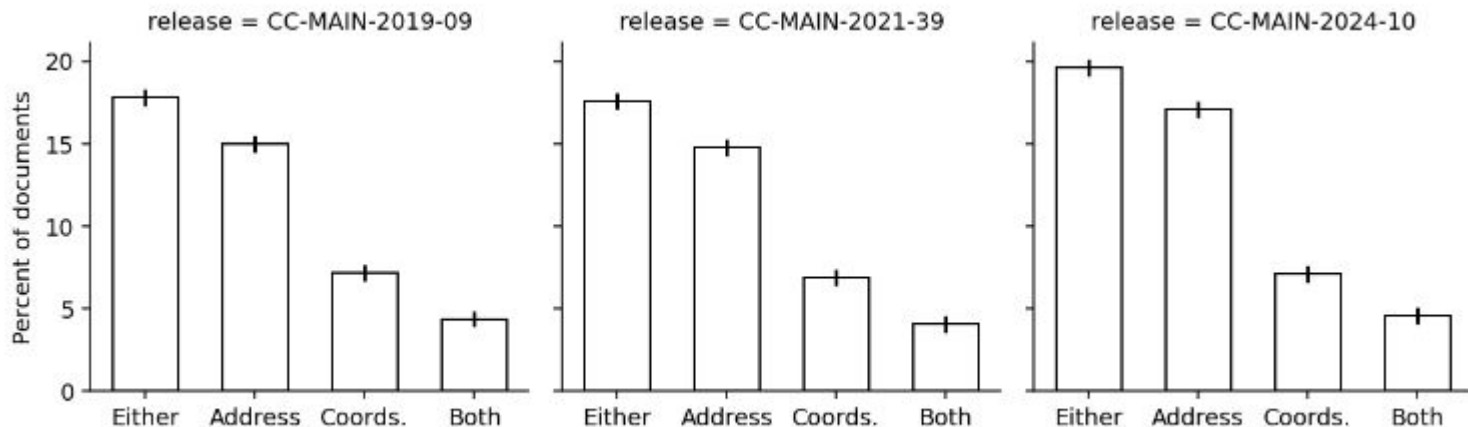


Figure 2: The prevalence of geospatial data in select CC releases, estimated within  $\pm 0.5\%$  at 95% confidence

Référence: Pre-print: Quantifying Geospatial in the **Common Crawl** Corpus  
- Ilyankou et al 2024. <http://arxiv.org/abs/2406.04952>

# Indicateur 3



## Waldo Tobler



Waldo Tobler en 2007.

---

**Waldo Tobler**, né en 1930 à [Portland](#)<sup>1</sup> et mort le 20 février 2018, est un cartographe et géographe [américano-suisse](#) [\[réf. nécessaire\]](#). Il est considéré comme un pionnier de la cartographie assistée par ordinateur et a développé plusieurs modèles de [cartogramme](#)<sup>2</sup>.

Il est aussi l'auteur de la « première loi de la géographie » selon laquelle « Tout interagit avec tout, mais deux objets proches ont plus de chances de le faire que deux objets éloignés »<sup>3</sup>.

Il est l'inventeur de la [projection hyperelliptique de Tobler](#)<sup>4</sup>.

Référence: Tobler - 1970 - A Computer Movie Simulating Urban Growth in the Detroit Region



# tetis-nlp/geographical-biases-in-llms

00.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text Cop

## Geographical

This tutorial aims to id

1. Spatial disparities
2. Spatial informati
3. Correlation betwe
4. Anomaly between

## Authors

Author	A
Rémy Decoupes	INR
Mathieu Roche	CIR
Maguelonne Teisseire	INR



Run all Ctrl+F9

Run before Ctrl+F8

Run the focused cell Ctrl+Enter

Run selection Ctrl+Shift+Enter

Run after Ctrl+F10

Interrupt execution Ctrl+M

Restart session Ctrl+M

Restart session and run all

Disconnect and delete runtime

Change runtime type

Manage sessions

View resources

View runtime logs

Pour la section 3.2 :

Ce notebook a besoin d'une GPU d'au moins 28 GB de RAM (GPU payante) :

- A100 (40BG)

Les sections 3.1 et 3.3 n'ont pas besoin d'environnement payant



# tetis-nlp/geographical-biases-in-llms

Distance between Taipei and Seoul: 1480.973652900838 km  
Distance between Taipei and Hanoi: 1669.0557310016254 km  
Distance between Taipei and Tokyo: 2104.49710309371 km  
Distance between Taipei and Singapore: 3240.6438418256876 km  
Distance between Taipei and London: 9803.291481023301 km

Similarity between Taipei and Seoul:	[[0.94581884]]	[[0.6629474]]	[[0.83839009]]
Similarity between Taipei and Hanoi:	[[0.94002306]]	[[0.61929286]]	[[0.84440932]]
Similarity between Taipei and Tokyo:	[[0.9552131]]	[[0.6553747]]	[[0.86354661]]
Similarity between Taipei and Singapore:	[[0.9569433]]	[[0.6196618]]	[[0.83802626]]
Similarity between Taipei and London:	[[0.9342403]]	[[0.59408295]]	[[0.80025356]]

RoBERTa

MISTRAL

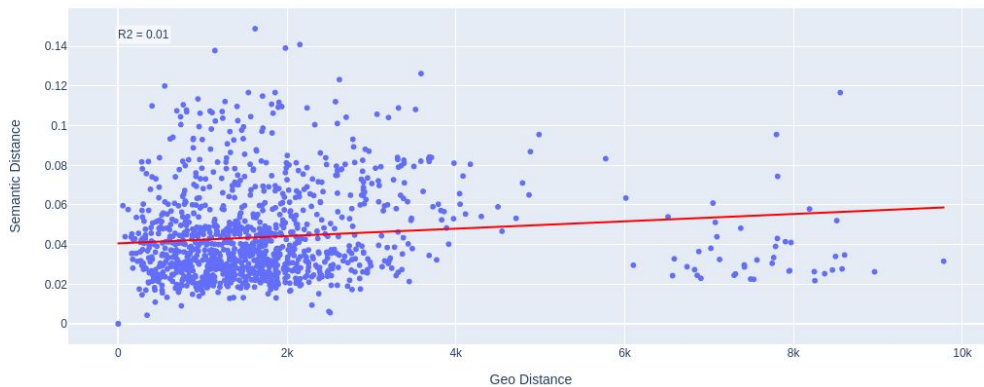
GPT-3.5

=> Tokyo trop proche sémantiquement de Taipei

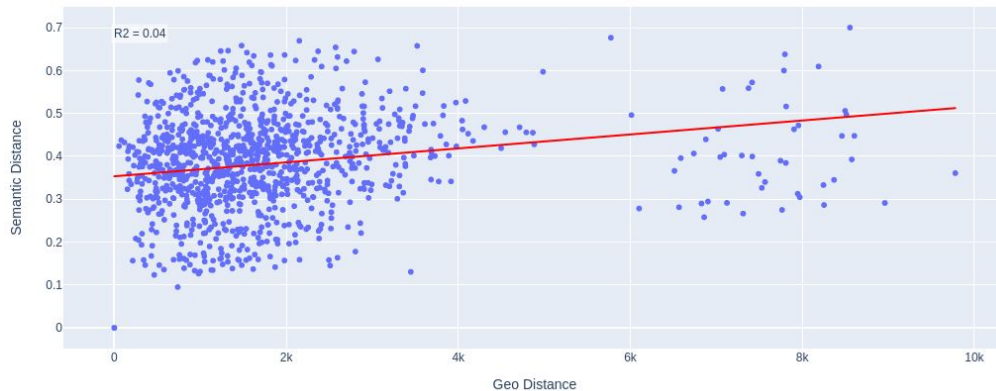


# tetis-nlp/geographical-biases-in-llms

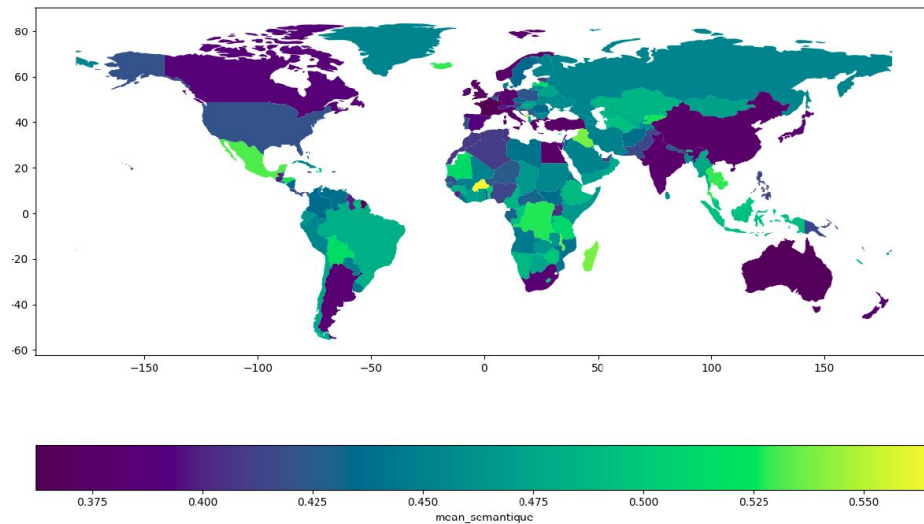
BERT :  
R2 : 0.01



GPT-3.5 :  
R2 : 0.04



# Indicateur 4





# tetis-nlp/geographical-biases-in-llms



00.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text Cop

Run all Ctrl+F9

Run before Ctrl+F8

Run the focused cell Ctrl+Enter

Run selection Ctrl+Shift+Enter

Run after Ctrl+F10

Interrupt execution Ctrl+M

Restart session Ctrl+M

Restart session and run all

Disconnect and delete runtime

Change runtime type

Manage sessions

View resources

View runtime logs

## Geographical

This tutorial aims to ide

1. Spatial disparities
2. Spatial informati
3. Correlation betwe
4. Anomaly between

## Authors

Author	A
Rémy Decoupes	INR
Mathieu Roche	CIR
Maguelonne Teisseire	INR



## Models

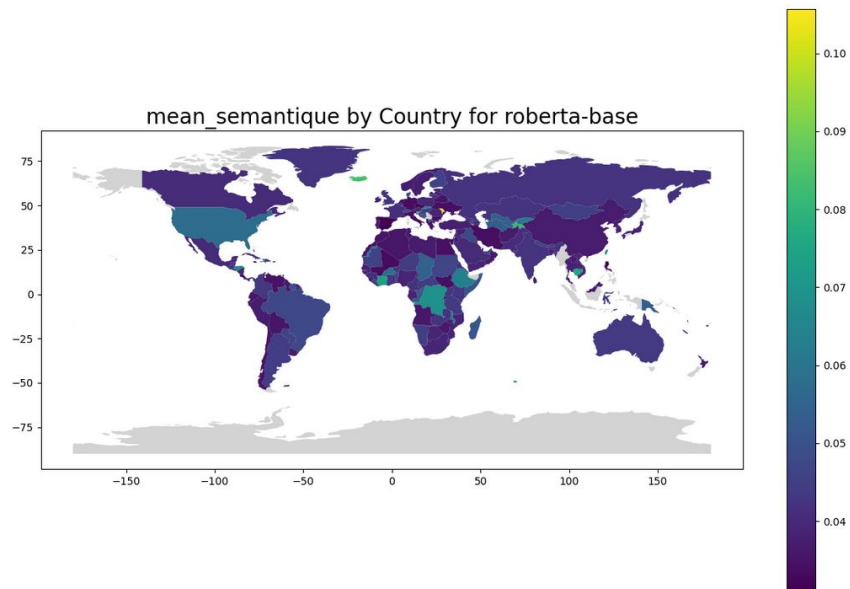
On peut rester sur un CPU !

**Pas besoin de changer  
d'environnement**

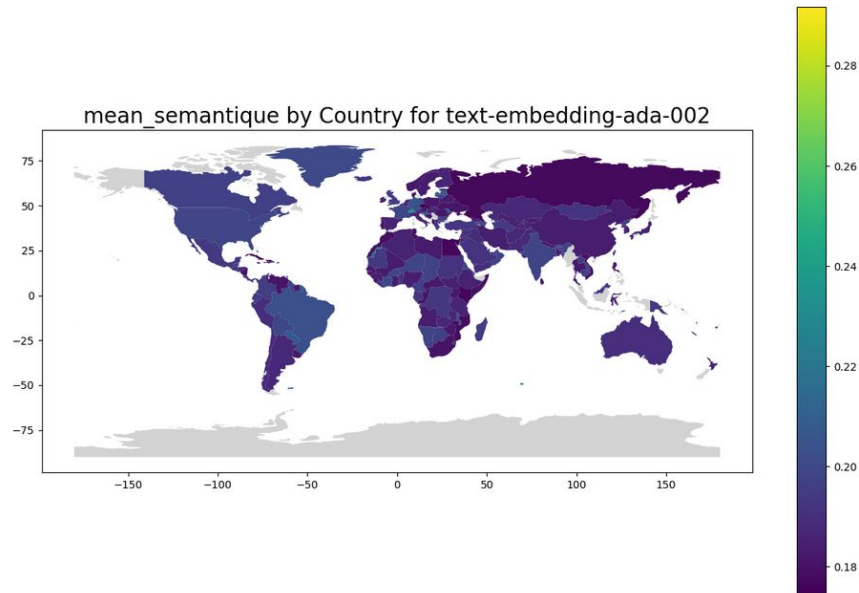




# tetis-nlp/geographical-biases-in-llms



RoBERTa



GPT-3.5

Aller plus loin



# tetis-nlp/geographical-biases-in-llms

1. Prédire le pays en fonction de sa capitale
  - a. Ajouter d'indicateurs (Population, Revenu / habitant / nombre d'hôpitaux / ...)
  - b. Améliorer les prompts pour améliorer la reproductibilité
2. Vocabulaire :
  - a. Comment évaluer indirectement la proportion d'info geo dans les corpus d'entraînement
3. Correlation distance Sémantique et Géographique
  - a. Clusterisation des pays en fonction de leurs proximité
4. Distorsion
  - a. Data visualisation : Quelles villes sont au centre de l'espace sémantique

***MERCI !***

**Session pratique : zoom sur les biais géographiques**



**Aurélie Névéol, Mathieu Roche, Rémy Decoupes**

LISN - Paris

UMR TETIS - Montpellier