

Practical session on Text Mining: from BERT to ChatGPT

Remy Decoupes & Maguelonne Teisseire

How recent concepts of NLP could be used in Epidemiological surveillance ?



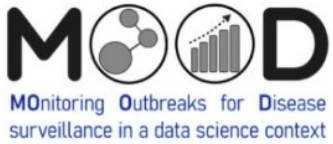


co nlp_pratical_session_BERT.ipynb

File Edit View Insert Runtime Tools Help [Changes will not be saved](#)

+ Code + Text Copy to Drive Connect

NLP for One Health

From BERT to ChatGPT

Speakers:

- Rémy DECOUPES - Research engineer UMR TETIS / INRAE
- Maguelonne TEISSEIRE - Prof. UMR TETIS / INRAE

Chapter 1. BERT

"[Bidirectional Encoder Representations from Transformers - Devlin et al - 2018](#)" from Google Research is an open-source pre-trained Language Model. BERT implements the well known "[Attention is all you need - Vaswani et al - 2017](#)"

Bert-case was trained on:

- Wikipedia (2.5 Billions of tokens)
- Google books (0.8 Billions of tokens).

On two tasks:

- Self-masking
- Next sentence prediction

1.1 Introduction: Transformers

A python library to easily work with BERT-like models

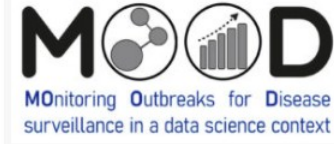


co nlp_pratical_session_ChatGPT.ipynb

File Edit View Insert Runtime Tools Help [Changes will not be saved](#)

+ Code + Text Copy to Drive Connect

NLP for One Health

From BERT to ChatGPT

Speakers:

- Rémy DECOUPES - Research engineer UMR TETIS / INRAE
- Maguelonne TEISSEIRE - Prof. UMR TETIS / INRAE

Chapter 2: ChatGPT

[ChatGPT](#) is an AI chatbot developed by OpenAI. ChatGPT uses a generative language model called GPT-3.5 without subscription and GPT-4 for 20€/month.

2.1 Requirements

As we are going to use ChatGPT through the [OpenAI API](#), participants need to create a account in order to obtain an API key: [here](#)

- Free access: 18\$ will be credited for a short period of time and it would not be renewed.
- Or add a payment method

Then let's begin with GPT3.5-turbo

```
[ ] # installation
!pip install openai
```