

Ilyes Boukaffal  
Pedro Alves  
L1-INT4  
2023-2024



# PYCHATBOT

Boukaffal-Alves-int4



# TABLE DES MATIÈRES :

- INTRODUCTION
- FUNCTIONAL PRESENTATION OF THE PROJECT
- TECHNICAL PRESENTATION OF THE PROJECT
- PRESENTATION OF THE RESULTS
- CONCLUSION

## INTRODUCTION :

The aim of this python project is to create a chatbot which, when given texts, must be able to analyze the texts, and be able to send us back answers, either through questions already recorded in the code, or through questions that the user asks and that the chatbot must in turn analyze.

The main objectives of this project are to help us develop our use of and skills in the Python language, and to solve problems by providing the clearest and most effective solutions possible. This project has enabled us to develop our skills in the use of functions, lists, arrays and dictionaries. The project has also encouraged us to look for information on our problems, whether in books, courses, blogs, etc.

## FUNCTIONAL PRESENTATION OF THE PROJECT :

- be able to ask questions that have already been recorded in the code.
- We can tell it the question we want to ask using the python console interface.

- The second feature: ask the chatbot any question, and it returns the file that best matches our question, as well as the sentence that best matches your question in that file.
- Unfortunately, this functionality was not completed, due to a number of problems that arose during the development of this part.

## TECHNICAL PRESENTATION OF THE PROJECT :

- The programme performs the various functions proposed by means of :
  - 
  - A TF-IDF matrix calculation, which calculates, among other things, the frequency of a word in relation to the frequency of a given document.
  - 
  - A vector similarity calculation, which enables the user to ask the chatbot any question they wish.
  - 
  - The choice of using dictionaries in our code is important, because it allows us to get closer to the architecture of a matrix, so that we can make full use of the way our TF-IDF calculation works.
  - 
  - The biggest technical problem was using GitHub, because there were a lot of file conflicts between participants.
  - 
  - The second problem was the display of special characters in matrices, which unfortunately has not been resolved, as well as a bug in the display of the answer to question 2.

## PRESENTATION OF RESULTS :

- User interface :

```
--- Menu ---

- To display the least important words, enter 1
- To display the most important words, enter 2
- To display the most important word by Chirac, enter 3
- To display the presidents who talk about the Nation, as well as the one who talks about it the most, enter 4
- To display the president who speaks most about ecology, enter 5
- Sorry, this command is not available at the moment : (To access the bot, enter 0.)

Enter a number between 0 and 5 :
```

- Matrix display :

|   |   |
|---|---|
| ['efforcera', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]      | ['exprimÃ©', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]     |
| ['fondÃ©', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]         | ['fortement', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]    |
| ['rÃ©els', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]         | ['durant', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]       |
| ['irrÃ©prochable', 0, 0, 0, 0, 0, 0, 0, 0]                  | ['achever', 0, 0, 0, 0, 0, 0, 0, 0]                       |
| ['battrai', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]        | ['confiÃ©', 0, 0, 0, 0, 0, 0, 0, 1.8061799739838869]      |
| ['mÃ©diterranÃ©', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]  | ['dÃ©cevoir', 0, 0, 0, 0, 0, 0, 0, 0]                     |
| ['afrique', 0, 0, 0, 0, 0, 0, 0, 0]                         | ['lorsqu', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]       |
| ['lutte', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]          | ['auxquels', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]     |
| ['rÃ©chauffement', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434] | ['confrontÃ©', 0, 0, 0, 0, 0, 0, 0, 0]                    |
| ['diplomatie', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]     | ['respecter', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]    |
| ['inscrire', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]       | ['donnÃ©', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]       |
| ['conviction', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]     | ['tenir', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]        |
| ['camp', 0, 0, 0, 0, 0, 0, 0, 0]                            | ['Ã©branlÃ©', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]    |
| ['bonnes', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]         | ['fragile', 0, 0, 0, 0, 0, 0, 0, 0]                       |
| ['volontÃ©s', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]      | ['morale', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]       |
| ['aiment', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]         | ['fort', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]         |
| ['animÃ©s', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]        | ['rÃ©habiliter', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434] |
| ['passion', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]        | ['tolÃ©rance', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]   |
| ['veulent', 0, 0, 0, 0, 0, 0, 0, 1.8061799739838869]        | ['intolÃ©rance', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434] |
| ['prÃ©t', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]          | ['sectarisme', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]   |
| ['travailler', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434]     | ['destructeurs', 0, 0, 0, 0, 0, 0, 0, 0.9030899869919434] |

- Display of answers for question 1 :

```
Enter a number between 0 and 5 : 1

- Voici les mots les moins importants

messieurs
les
mesdames
en
ce
je
la
de
le
une
a
france
qui
se
et
dans
peuple
```

- Display of answers for question 2 :

```
Enter a number between 0 and 5 : 2  
  
- Voici le mot le plus important du dossier :
```

- Display of answers for question 3 :

```
Enter a number between 0 and 5 : 3  
  
- Voici le mot le plus important des discours de Chirac : doit
```

- Display of answers for question 4 :

```
Enter a number between 0 and 5 : 4  
  
- Voici les présidents parlant de Nation  
  
Chirac  
Hollande  
Macron  
Mitterrand  
  
- Mitterrand parle le plus de Nation
```

- Display of answers for question 5 :

```
Enter a number between 0 and 5 : 5  
  
- Macron est le président qui parle le plus du climat
```

## CONCLUSION :

In terms of soft skills, this project has enabled us to learn about group work, how to organize ourselves and divide tasks, how to manage our time well between different academic projects, and how to find the right information where it's needed: on the internet, in books, in courses, etc.

In terms of hard skills, this project enabled us to strengthen our skills in Python, such as the use of functions, 1D and 2D lists and dictionaries. It also enabled us

to learn more about the procedural paradigm, as at lycée we focus more on object-oriented programming.