

Challenge – NLP Engineer

Introduction

Congratulations for making it to the Challenge stage of the application process! The goal of this task is to objectively assess your technical prowess and your ability to solve business problems using state of the art NLP techniques, database and potentially custom machine learning.

Please note that the topic, data and problem are reflective of ongoing cases.

Deadline: 7 Days

Deliverable:

- A .zip file with contents grouped into the following sub-directories (you may omit empty directories):
 - o Code (all the files and codes you would generate to solve the challenge)
 - o Slides (mainly a presentation of your work and any other documents that you want to highlights)
- Please label the .zip file: <your_first_name>_<your_last_name>_nlp.zip, for example: karim_daoui_nlp.zip

The Task

We are building a state of the art conversational chatbot that uses financial documents covering trades. The objectif of the chatbot is to answer user questions with the highest accuracy (stay true to the documents).

More about conversational AI : [Conversational AI in 2023: What It Is, How It Works & Use Cases \(aimultiple.com\)](https://aimultiple.com)

You have at hand two pdf files that describes risk factors related to a financial trade (two documents each for its own trade), the user of our chatbot want to get answers based on the knowledge of the documents.

example

user: show me an outline of the rsik factors

chatbot: the risk factors outlined in the first document are:

1. ...
2. ...
 - a. ...
 - b. ...
3. ...

And the risk factors outlined in the second document are:

1. ...
2. ...
3. ...
 - a. ...
 - b. ...
 - c. ...
 - d. ...
4. ...
 - a. ...
 - b. ...
 - i. ...
5. ...

To help you get started use this questions to kick start your journey.

1. What technology are you going to choose to build the socle of the chatot?
2. What techniques are you going to choose to make the model?
3. Now, let's assume that you have the langage model that will allow you to answer the user questions.
 - a. What is the metric that are you going to choose to evaluate it
 - b. What dataset are you going to use (probably you should build one)
4. At a certain point you will need a database to store the documents (as we will add more documents: here only 2 at this challenge):
 - a. What is this database?
 - b. Please explain in details the benefits of your choice compared to other alternatives.

Documents

- offering memorandum
- consular insurance

both documents are similar in structure, yet they have dissimilarities in content.

Both describes risk factors related to financial trades

Tips

1. Explore state of the art solutions and what is happening right now in the field
 - a. openAI
 - b. LlamaIndex
 - c. Langchain
 - d. ...etc
2. Show us your coding skills
3. Communicate your thought process
4. Play to your strenght – spend more time in the model, database, conversational AI or the retrieval / search depending on your skills