

**Stage de 5 mois**  
**avril 2023 - août 2023**  
**Application d'orientation au sein du Telegram**

*Formation : Master II en Sciences et Technologies du Logiciel (STL)*

*Employeur : l'association*

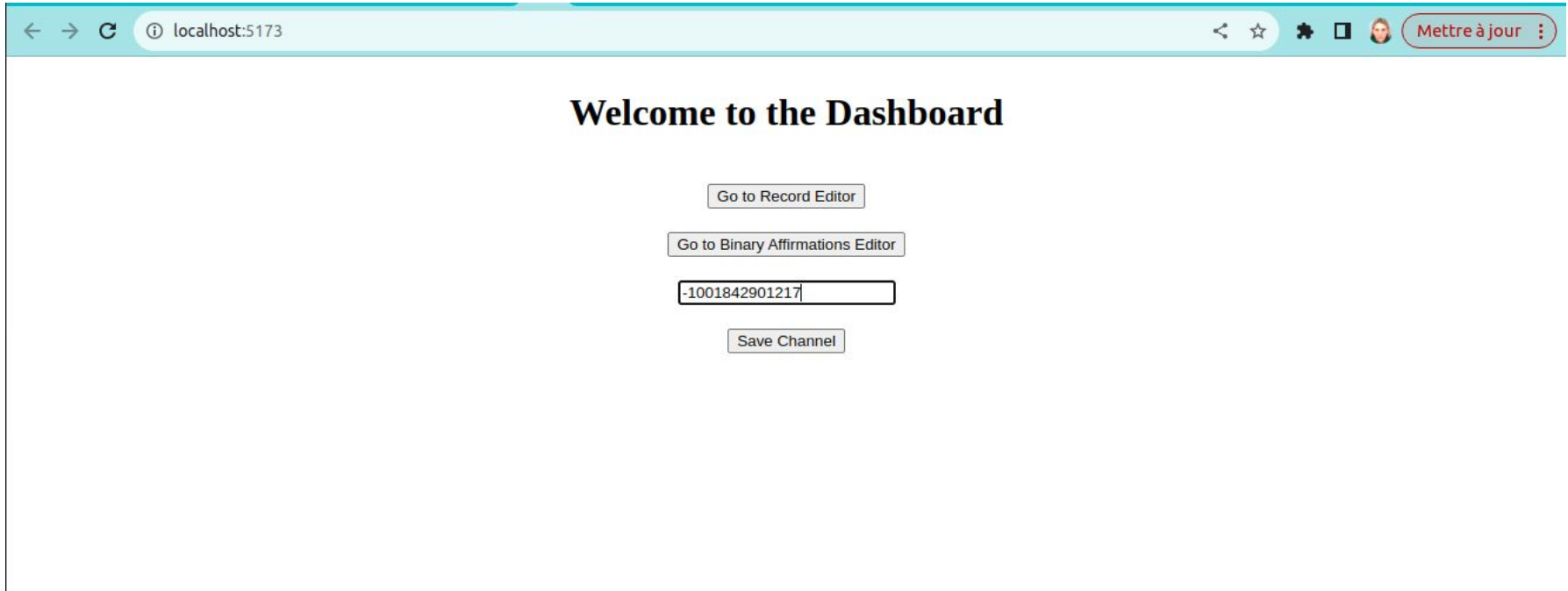
*Objectif : - aborder la vérification des informations contenues dans le Telegram*  
*- valoriser les compétences et en acquérir de nouvelles*  
*- se lancer dans l'informatique à vocation sociale*

# Plan

1. Vu utilisateur
2. Développement
3. Tests
4. Des limites
5. Conclusion

# 1. Vu utilisateur

Un utilisateur tape les noms des chaînes



The screenshot shows a web browser window with the address bar displaying 'localhost:5173'. The browser's toolbar includes navigation icons (back, forward, refresh), a search icon, a star icon, a puzzle piece icon, a square icon, a user profile icon, and a red button labeled 'Mettre à jour' with a dropdown arrow. The main content area of the browser displays a dashboard with the heading 'Welcome to the Dashboard'. Below the heading, there are three buttons: 'Go to Record Editor', 'Go to Binary Affirmations Editor', and 'Save Channel'. A text input field is positioned between the 'Go to Binary Affirmations Editor' and 'Save Channel' buttons, containing the text '-1001842901217'.

← → ↻ ⓘ localhost:5173 🔍 ☆ 🧩 ◻ 👤 Mettre à jour ⋮

## Welcome to the Dashboard

Go to Record Editor

Go to Binary Affirmations Editor

-1001842901217

Save Channel

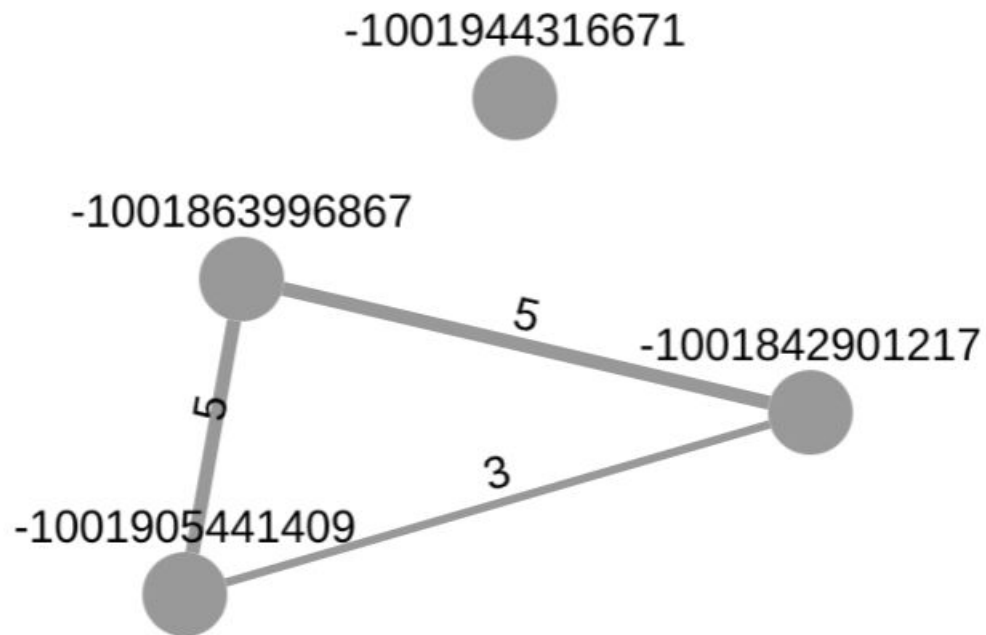
## 1. Vu utilisateur

Il obtient les informations suivantes par rapport à ces chaînes :

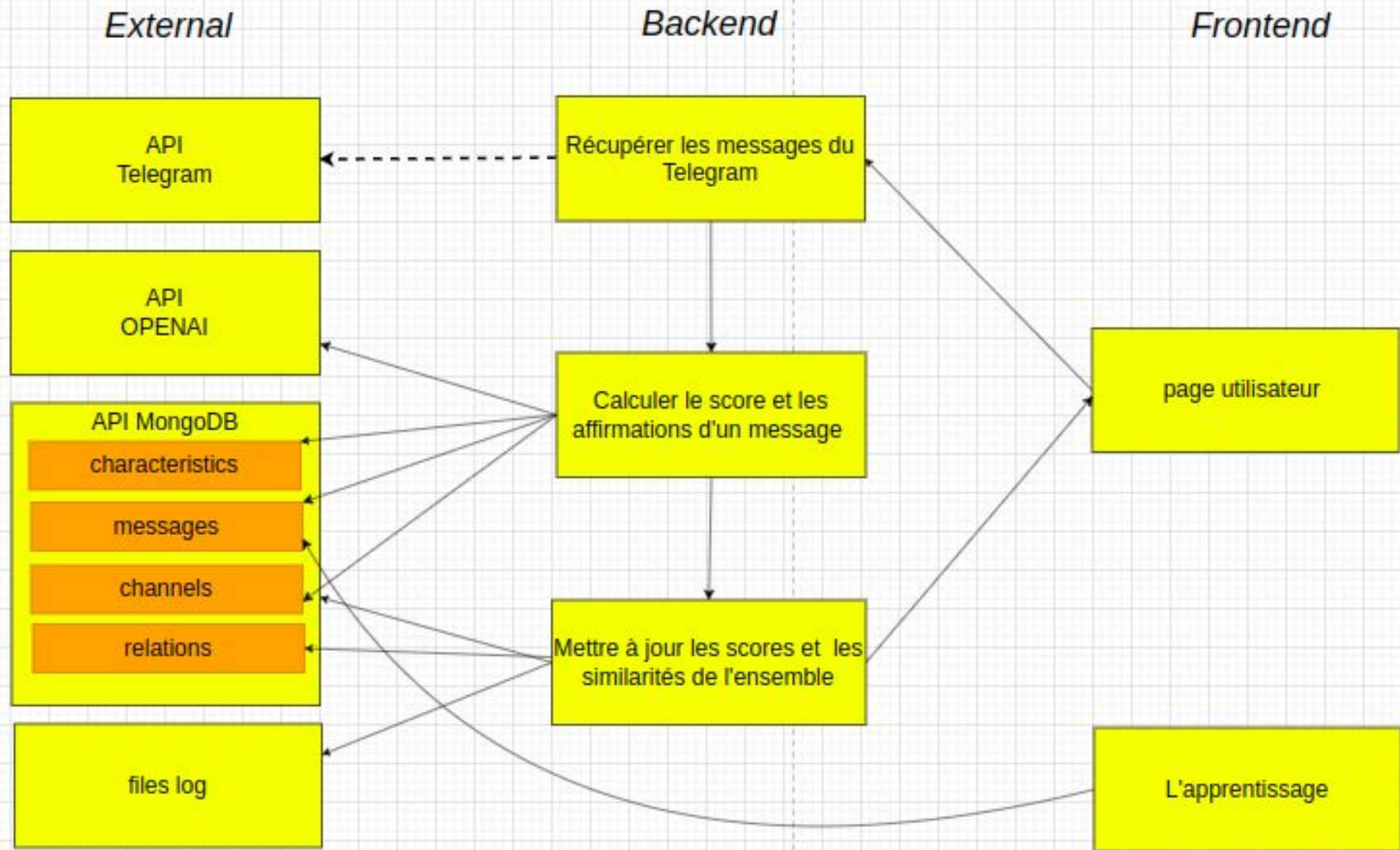
- 1) la mesure de similarité entre les chaînes deux par deux  
similarité (channel\_1, channel\_2) = nombre d'affirmations similaires  
similarité (channel\_1, channel\_2)  $\in [0 \dots N]$   
 $N = \min (\text{nombre d'affirmations channel\_1}, \text{nombre d'affirmations channel\_2})$
- 2) le coefficient de confiance pour chaque chaîne  $[0 \dots 100]$

Ces informations sont affichées sous la forme d'un graphe.

## 1. Vu utilisateur



## 2. Développement : Architecture



## 2. Développement : Les paramètres (OpenAI)

temperature

max\_tokens

max\_len\_message

prompt\_c : récupérer les caractéristiques d'un message

26 questions (trouvées dans la littérature spécialisée, et ajoutées par les testes)

récupérer les caractéristiques d'un message (une chaîne, un groupe de chaînes)

prompt\_a : récupérer l'essentiel d'un message

## 2. Développement : Récupérer les caractéristiques d'un message

*Analyze the following message and provide its characteristics. Reply by numbers, one number for each criterion and nothing else.*

*<message>*

*1. positive author attitude: {if true reply -1, if false reply 0, if neutral reply 0}*

*2. the message is advantageous for the actual political power: {if true reply -1, if false reply 0, if neutral reply 0}*

*...*

*27. underestimation of the proofs of the opponents: {if true reply -1, if false reply 0, if neutral reply 0}*



## 2. Développem. : Récupérer les affirmations (l'essentiel) d'un message

*Please generate binary affirmations from the provided news message:*

*'{message}'*

*For each affirmation:*

*Ensure that each binary affirmation is concise, clear, and independent. They should almost function as a unique hash of the information they contain.*

*For example. Here I provide examples already checked by gpt-4, with corrections = result that should appear instead, or close to it. You need understand what I mean with these examples and strictly follow :    """"*

## 2. Développement : Exemple des affirmations

La France va connaître des pointes de chaleur à 35°C entre dimanche et lundi. Lundi pourrait être la plus chaude journée d'octobre jamais mesurée en France.

Message:

### GPT Extracted Affirmations:

France will experience heat peaks of 35°C between Sunday and Monday

true

Monday could be the hottest October day ever recorded in France

true

## 2. Développement : Apprentissage caractéristiques

Characteristics	Score	Corrected Score	Explanation
positive author attitude	-1	0	no positive attitude
the message is advantageous for the actual political power	0	0	
exaggerations	0	0	
lack of object concordance	0	0	
appeal to the intellect	0	0	
appeal to emotions	0	0	
appeal to virtues	0	0	
appeal to ethics	0	0	
appeal to blinding generalities as fatherland, peace, glory, justice, honor	0	0	
appeal to fear	0	0	
appeal to the threat from other countries	0	0	
appeal to hatred	0	0	
appeal to authority	0	0	
appeal to human rights	0	1	приоритетам о гражданском
criticism of the actual political power	0	0	
generalizations	0	0	
seemingly scientific appearance, but only seemingly	0	0	
real scientific proofs of the information	0	0	
real proofs of the information	0	0	
euphemismes replacing offensive expressions	0	0	
euphemismes replacing war or victims	3	0	there is no such euphemismes
formal register of the texte	0	-1	it is formal
eupbeneficiaries of the information regarding to politics	0	0	
flattery to the reader	0	0	
dilemma between only two possibilities	0	0	

## 2. Développement : Apprentissage les affirmations

### Edit Binary Affirmation for ID: 651f49481f265178245b8d99

Message: \*\*В ЕС обеспокоены процессом принятия в

#### GPT Extracted Affirmations:

The EU is concerned about the adoption	true
Brussels notes that the final adoption	true
The law contradicts key priorities about	true

#### Corrected Affirmations:

The EU is concerned about the adoption	true
Brussels notes that the final adoption	true
The law contradicts key priorities about	true

☐ Use this affirmation

Save Back

## 2. Développement : Complexité

$O(N)$  requêtes OpenAI

$O(N * K)$  requêtes MongoDB

N nombre total de messages

K nombre chaînes

## 2. Développement : Détails

Classes: Group, Message, Prompt\_c, Prompt\_a, Characteristic

## **2. Développement : Les changements par rapport au rapport**

- 1) une seule application à la place de deux
- 2) un seul API public pour lancer l'application et l'apprentissage
- 3) pas d'apprentissage niveau tchat
- 4) les requêtes OpenAI ne sont asynchrones, les requêtes MongoDB ne sont pas toutes asynchrones
- 5) formule

### 3. Tests : Les 26 questions

SET 1 \*\*\*\*\* 81

@eschulmann 84/100  
@teamnavalny 82/100  
@nevzorovtv 77/100  
@cherniy\_kvartal 82/100  
@khodorkovski 83/100  
@stalin\_gulag 80/100  
@viktor\_shenderovich\_izbrannoe 80/100

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
0.0	-0.1	-0.0	0.0	0.1	-0.1	0.0	0.0	-0.0	-0.0	-0.0	-0.1	-0.0	0.2	0.1	-0.0	0.0	0.0	0.0	0.0	-0.0	-0.0	-0.1	0.0	0.0	-0.1	-0.0

SET 2 \*\*\*\*\* 78

@SolovievVRNews 81/100  
@vv\_volodin 79/100  
@MariaVladimirovnaZakharova 84/100  
@medvedev\_telegram 77/100  
@donbassinsider 82/100  
@RKadyrov\_95 67/100

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
0.0	-0.2	-0.0	0.0	0.0	-0.2	-0.0	-0.0	-0.0	-0.0	-0.1	-0.1	-0.1	0.3	0.1	-0.1	-0.0	0.0	0.0	0.0	0.0	-0.0	-0.2	-0.0	0.0	-0.1	-0.0

!



### 3. Tests : Temperature

Temperature	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
l'écart	2	2	2	2	3	4	3	3	5	5	5
l'écart	1	1	2	3	1	1	3	4	2	4	4

## 4. Limites : Limites techniques liées à OpenAI

- 1) Les défauts d'analyse par OpenAI
- 2) Manque de scalabilité (GPT-3 20 messages par minute, GPT-4 5)
- 3) Des fois connexion à API OpenAI échouée
- 4) max\_len\_message
- 5) Les affirmations - seulement GPT-4
- 6) Peu de testes GPR-4

## 4. Limites : Autres limites techniques

6) des chaînes ne se lisent pas (<https://t.me/generallsvr>)

7) les 26 questions testées principalement sur un ensemble de 13 chaînes

## **4. Limités : Non techniques**

8) payant

9) l'utilité discutable

10) le choix des caractéristiques et l'apprentissage reposent sur une opinion subjective

## 5. Conclusion

- 1) Des choses apprises
- 2) Améliorations de ce projet possibles :
  - la scalabilité
  - d'alerte pour des situations anormales
  - la première chaîne à diffuser une information
  - le traitement croisé par plusieurs AI
  - la comparaison des résultats avec d'autres projets
- 3) D'autres projets