



Paris
School of
Technology
& Business

LLM FINE-TUNÉ POUR L'ANALYSE DE SENTIMENT AUTOUR DU CLIMAT ET LES RÉPONSES CONTEXTUELLES

29 Juillet 2025 - Team BAD GETAWAY - Karim Ben Yahia & Léa Mariani

PROBLÉMATIQUE ET OBJECTIF

Les entreprises croulent sous les messages clients (réseaux sociaux, forums, emails). Pourtant, elles peinent à capter rapidement le **sentiment** et à produire une **réponse contextualisée**.

L'objectif est de construire un pipeline modulaire capable de faire les deux, en s'appuyant sur des modèles de langage fine-tunés.

SUITE LOGIQUE

Lors du premier hackathon, on a pris de la hauteur pour capter les grandes tendances autour du climat. Cette fois, on entre dans le détail des émotions, des intentions, et des réponses possibles...

**...pour enfin passer de
l'analyse à l'ACTION !**



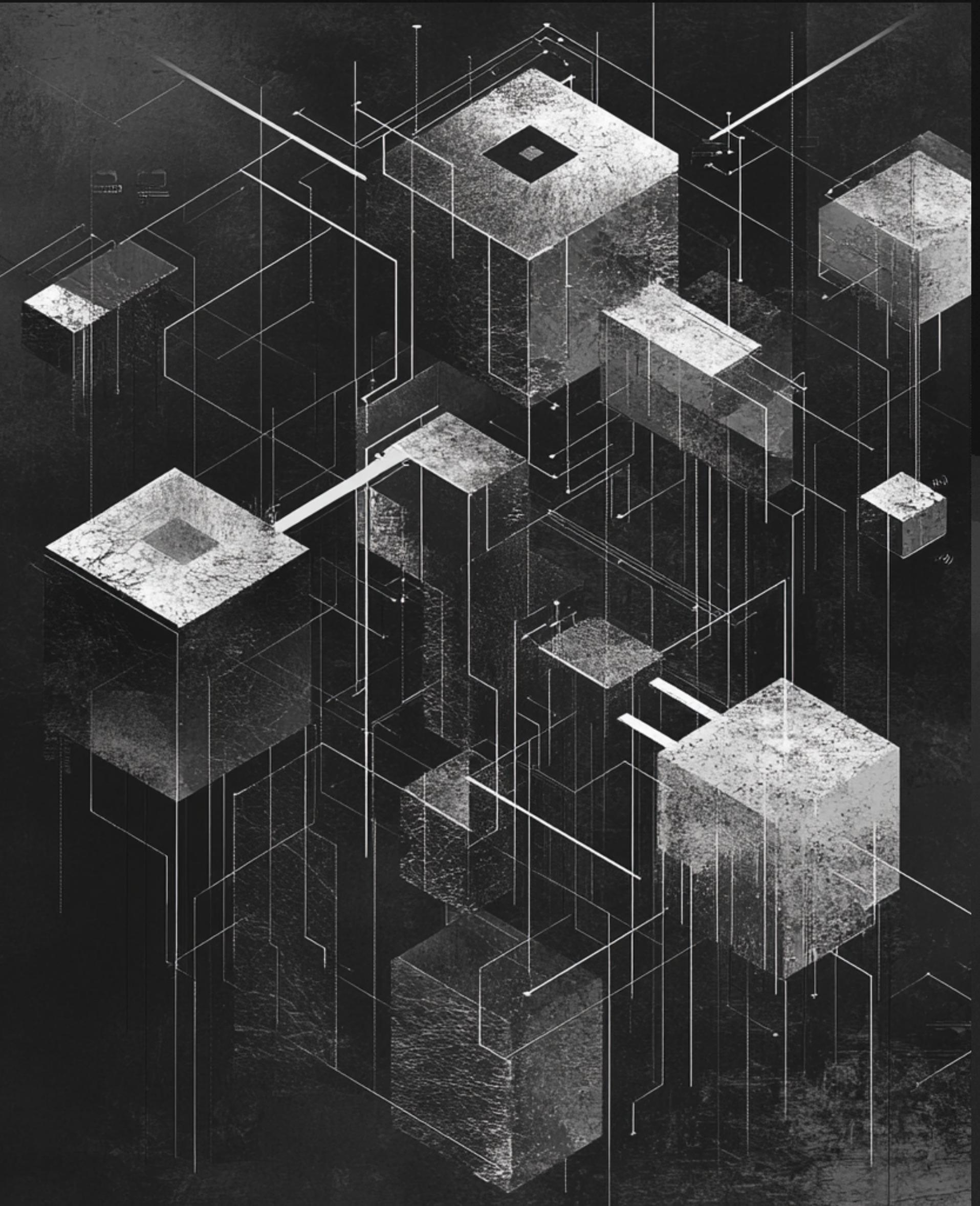
DÉFIS CLÉS

1. Choix d'un modèle performant : **LORA & DistilBert** (combinaison de modèles légers et adaptés au contexte).
2. Trouver une solution **compatible GPU & CPU**.
3. Mettre en place une **architecture efficace** pour la première fois.
4. Gérer les **compatibilités entre modules**.
5. Gérer les **compatibilités entre les librairies**.
6. Faire face aux **ressources limitées** en GPU.
7. Temps imparti restreint.



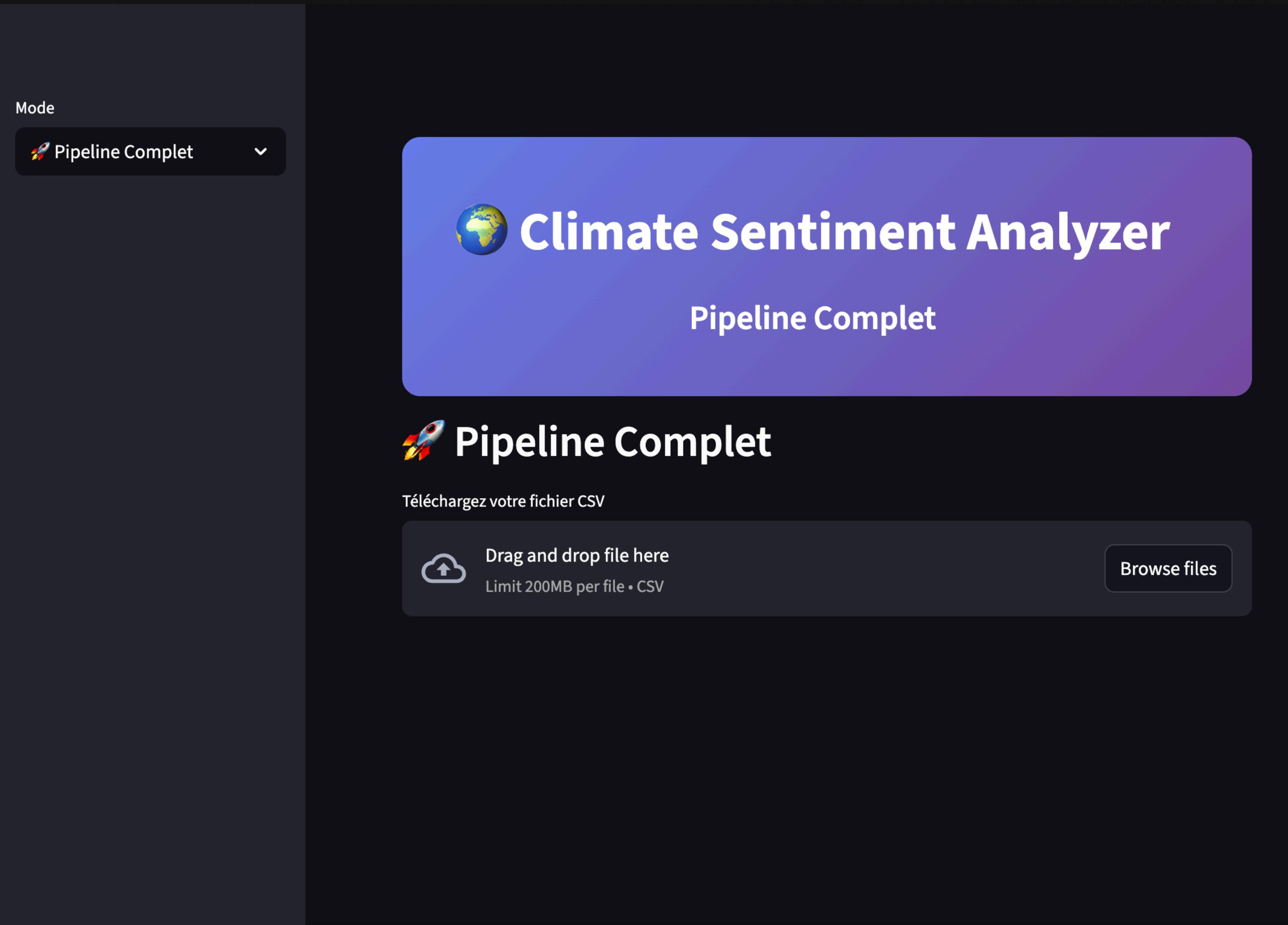
FLOWCHART

```
streamlit_app.py          # Interface principale  
core_modules.py          # Configuration centrale  
data_modules.py          # Chargement et nettoyage CSV  
model_modules.py         # Modèle + LoRA + Trainer  
visualization_modules.py # Graphiques et rapports  
qa_modules.py            # Recherche sémantique  
knowledge_modules.py     # Base de connaissances  
requirements.txt          # Dépendances  
README.md                # Ce fichier  
outputs/                 # Sauvegardes du modèle
```



DÉMO

streamlit_app.py # Interface principale



DÉMO

|--- **data_modules.py**

```
# Chargement et nettoyage CSV
```

The screenshot shows a web application interface for a "Climate Sentiment Analyzer". At the top left, there is a "Mode" dropdown menu set to "Pipeline Complet". The main title "Climate Sentiment Analyzer" is displayed prominently at the top right, along with a small globe icon.

The central area is titled "Pipeline Complet". It features a "Drag and drop file here" input field with a "Browse files" button. A file named "subsampled.csv" (3.3MB) is listed as uploaded. A green success message indicates that the file was successfully loaded with 10000 lines and 2 columns.

At the bottom, a table preview shows three rows of data:

| | self_text | comment_sentiment |
|---|--|-------------------|
| 0 | Comment has been removed for breaking Rule 1: > **Direct responses to a CMV po... | 0 |
| 1 | It's not even a secret, ask your average Russian about climate change, either they don'... | 1 |
| 2 | ... | ... |

DÉMO

|—— model_modules.py # Modèle + LoRA + Trainer

The screenshot shows a user interface for training a machine learning model. At the top, there's a file upload section with a CSV file named "subsampled.csv" (3.3MB). A green success message indicates the file contains 10000 lines and 2 columns. Below this is a table preview showing five rows of data with columns "self_text" and "comment_sentiment". Further down, there are two sliders for "Taille échantillon" (set at 4000) and "Epochs" (set at 3), both with scales from 1000 to 10000. A large red button at the bottom right says "Lancer l'entraînement" (Launch training).

Mode

Pipeline Complet

subsampled.csv 3.3MB

Fichier chargé : 10000 lignes, 2 colonnes

| | self_text | comment_sentiment |
|---|---|-------------------|
| 0 | Comment has been removed for breaking Rule 1: > **Direct responses to a CMV po: | 0 |
| 1 | It's not even a secret, ask your average Russian about climate change, either they don' | 1 |
| 2 | >It has not solved language processing problems and still cannot handle translation | 1 |
| 3 | No, I absolutely do know they are giant balls of gas. Surely you have some sort of proc | 1 |
| 4 | The paper here suggests a method to reduce total agricultural land by 75% or so. A lot | 1 |

Taille échantillon

4000

1000 10000

Epochs

3

1 5

Lancer l'entraînement

DÉMO

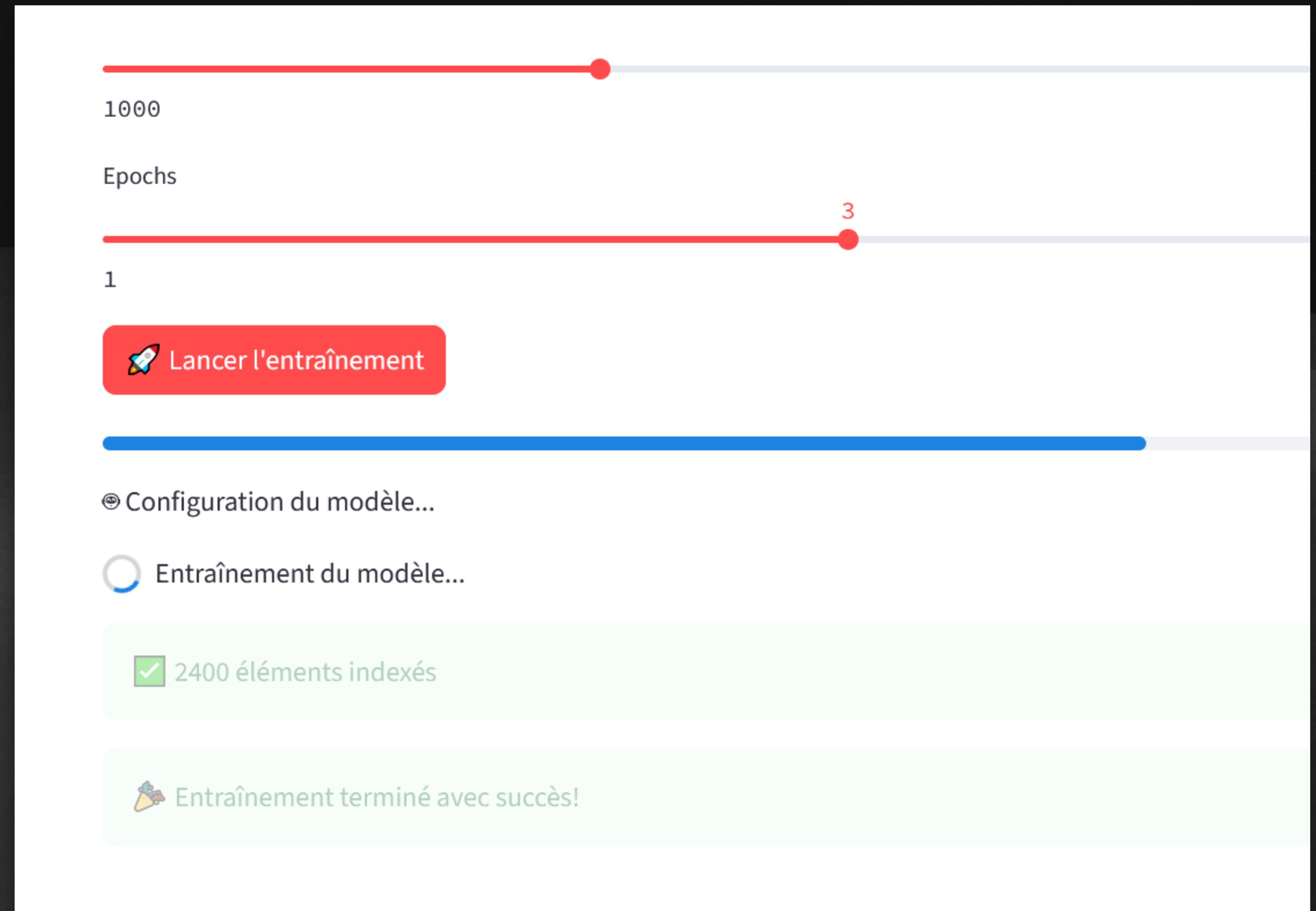
|—— model_modules.py # Modèle + LoRA + Trainer

The screenshot shows a machine learning training interface with the following elements:

- Mode:** Pipeline Complet
- Fichier chargé:** 10000 lignes, 2 colonnes
- Tableau de données:** Affiche 5 lignes d'un dataset avec les colonnes self_text et comment_sentiment.
- Taille échantillon:** 1000 (réglage)
- Epochs:** 1 (réglage)
- Bouton "Lancer l'entraînement":** Bouton rouge pour démarrer le processus.
- Progress bar:** Indique la progression de la tokenisation des données.
- Message vert:** Données préparées: Train=600, Val=200, Test=200

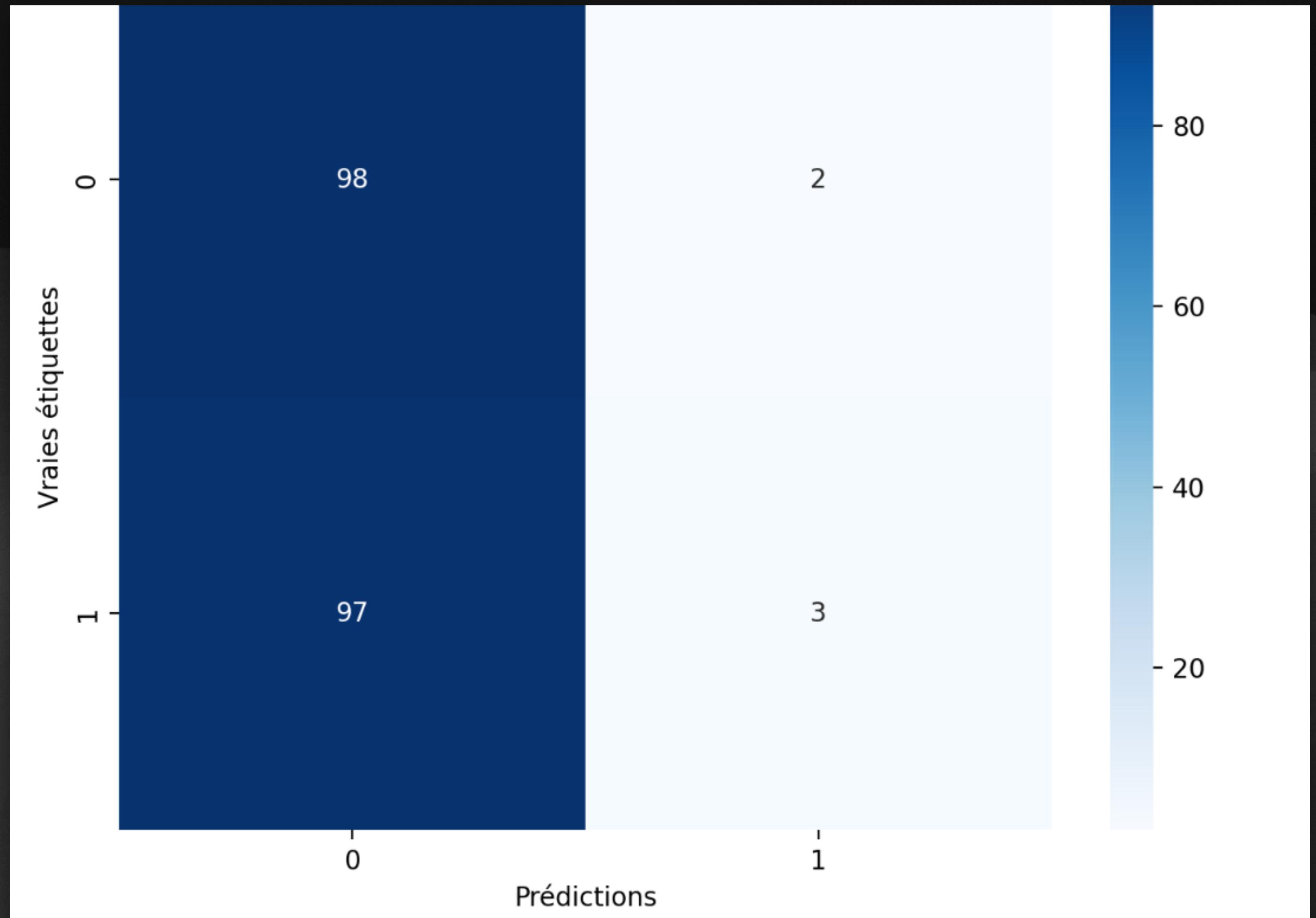
DÉMO

|—— model_modules.py # Modèle + LoRA + Trainer



DÉMO

└── **visualization_modules.py** # Graphiques et rapports



DÉMO

└── **visualization_modules.py** # Graphiques et rapports

Choisir le type de visualisation:

Rapport de classification ▾

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.5026 | 0.98 | 0.6644 | 100 |
| 1 | 0.6 | 0.03 | 0.0571 | 100 |
| accuracy | 0.505 | 0.505 | 0.505 | 0.505 |
| macro avg | 0.5513 | 0.505 | 0.3608 | 200 |
| weighted avg | 0.5513 | 0.505 | 0.3608 | 200 |

DÉMO

|--- qa_modules.py

Recherche sémantique



Climate Sentiment Analyzer

Pipeline Complet

? Interface Q&A

Votre question:

Ex: Quelles sont les causes du réchauffement climatique?

Ce projet offre un pipeline IA modulaire pour **analyser les sentiments climatiques**, combinant prétraitement automatisé, fine-tuning léger, visualisation, génération de réponses et interface interactive. **Facilement déployable** via Streamlit, il répond aux défis de données hétérogènes et d'interprétabilité grâce à des choix techniques efficaces comme

LoRA et **DistilBert**

BÉNÉFICES MÉTIER

/ Automatisation et industrialisation des workflows IA

Réduction du temps et des coûts de développement (pipelines standardisés).

Exécution fiable et répétable des traitements (préprocessing, entraînement, QA).

/ Amélioration de la qualité des modèles

Suivi et évaluation centralisée (module QA).

Meilleure explicabilité via la visualisation et les métriques.

/ Montée en productivité des équipes

Déploiement plus rapide grâce au module Streamlit (interface prête à l'emploi).

Réduction des tâches manuelles (gestion des données et configurations automatisées).

/ Scalabilité et modularité

Ajout facile de nouveaux modèles ou jeux de données.

Architecture modulaire permettant la réutilisation dans d'autres projets.

/ Prise de décision améliorée

Meilleure exploitation des données via l'intégration d'une knowledge base.

Accès rapide aux résultats et analyses via des visualisations interactives.





Paris
School of
Technology
& Business

ET SI DEMAIN, CES OUTILS DEVENAIENT DES ASSISTANTS DE DÉCISION POUR LA TRANSITION ÉCOLOGIQUE,
JUSQU'ΟÙ POURRIONS-NOUS ALLER COLLECTIVEMENT ?

Merci ! À vos questions

