

Technique de programmation

Thomas - Grégoire - Essi

Université de Strasbourg

Avril 2025

But du projet

- Création d'un modèle capable de répondre à des questions

On pose une question :

“Où trouver des légumes en circuits courts près de Meinau”

- Selon le prompt, le modèle nous oriente localement en fonction de la demande

Le web scraping

Pourquoi faire ?

- Données fiables et récentes
- Éviter les biais
- Adapter le modèle au territoire
- Centraliser l'information



Figure: Fiabilité

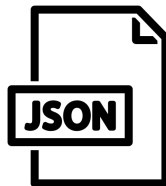


Figure:
Centralisation

Le web scraping

Comment ?

- Phase la plus longue : trouver des sites récents et pertinents
- On rassemble toutes les urls dans un fichier .txt
- On boucle sur le fichier, en webscrappant chaque site, puis on rassemble dans un fichier json la data, sous la forme d'un dictionnaire

```
with open('/content/urls.txt', "r") as f:
    urls = [line.strip() for line in f if line.strip()]

results = []

for url in urls:
    print(f"Scraping: {url}")
    downloaded = trafilatura.fetch_url(url)

    if downloaded is not None:
```

Figure: Boucle for

```
{
  "url": "https://www.reussir.fr/fr/...",
  "texte": "« Un concept innovant p...",
},
{
  "url": "https://www.dna.fr/politi...",
  "texte": "Gundershoffen Un vrai é...",
},
{
  "url": "https://www.dna.fr/societe...",
  "texte": "Gundershoffen L'épicerie..."
}
```

Figure: Structure dictionnaire

Le web scraping

Comment ?

- Web Scrapping de plusieurs sites
- Utilisation du package Trafilatura
- Cela évite d'avoir à utiliser du regex, pour chaque site



Figure: Trafilatura

```
# import the package
>>> import re

# syntax of findall function
>>> re.findall(pattern = <replace with the regex pattern>,
               string = <replace with the text>)

# example usage
>>> re.findall(pattern = r"Player\d",
               string = "Player1 and Player2 form a team.")
['Player1', 'Player2']
```

Figure: Regex

Le web scraping

API

- Utilisation de l'API data.strasbourg.eu
- Très grosse quantité d'information

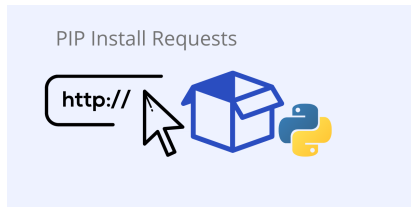


Figure: Requests

L'interface graphique

Les fonctionnalités

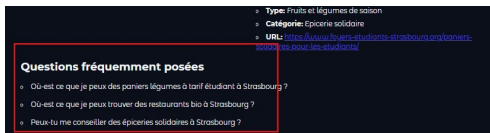
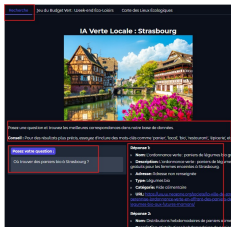
- Recherche intelligente d'informations écologiques locales
- Un jeu interactif pour comprendre l'impact carbone des loisirs
- Une carte interactive des lieux écologiques à Strasbourg

L'interface graphique

L'onglet recherche

L'interface de cet onglet est divisée en plusieurs sections claires pour guider les utilisateurs.

- la zone de consignes et de conseils : aide les utilisateurs à formuler leurs questions de manière précise en utilisant des mots-clés pertinents comme 'panier', 'local', 'bio', etc.
- Zone de saisie et Zone d'affichage de réponses
- Questions fréquemment posées: En prenant l'une de ces questions, les utilisateurs peuvent rapidement obtenir une réponse.

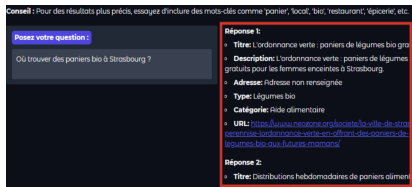


L'interface graphique

L'onglet recherche

Quand l'utilisateur fait une recherche, le système:

- Analyse puis identifie les termes clés utilisés, par exemple 'Localisation'...
- Rechercher des correspondances rapidement grâce à un index construit à partir de notre base de données en fichiers JSON.
- Calcule un "score de pertinence" pour les résultats trouvés
- Les résultats sont triés par score de pertinence, et seules les trois meilleures réponses sont conservées



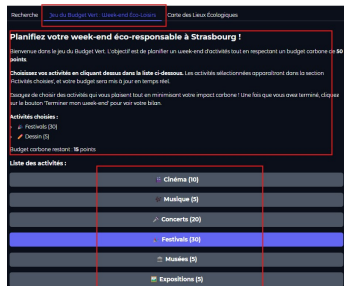
L'interface graphique

L'onglet "Jeu du Budget Vert : Week-end Éco-Loisirs"

Objectif du jeu

Planifier un week-end à Strasbourg en respectant un budget carbone

- Chaque activité est associée à un 'coût carbone'
- Bouton "Terminer mon week-end" : déclenche l'évaluation du bilan carbone de l'utilisateur



L'interface graphique

L'onglet "Jeu du Budget Vert : Week-end Éco-Loisirs"

Recherche Jeu du Budget Vert : Week-end Éco-Loisirs Carte des Lieux Écologiques

Planifiez votre week-end éco-responsable à Strasbourg !

Bienvenue dans le jeu du Budget Vert. L'objectif est de planifier un week-end d'activités tout en respectant un budget carbone de **50 points**.

Choisissez vos activités en cliquant dessus dans la liste ci-dessous. Les activités sélectionnées apparaîtront dans la section 'Activités choisies', et votre budget sera mis à jour en temps réel.

Essayez de choisir des activités qui vous plaisent tout en minimisant votre impact carbone ! Une fois que vous avez terminé, cliquez sur le bouton 'Terminer mon week-end' pour voir votre bilan.

Activités choisies :

- 🎪 Festivals (30)
- 🎨 Dessin (5)

Budget carbone restant: **15 points**

Liste des activités :

- 🎬 Cinéma (10)
- 🎵 Musique (5)

🏠 Vacances (45)

🎓 École (15)

💼 Travail (10)

Vos choix :

Terminer mon week-end

Bilan de votre week-end éco-responsable :

Votre budget initial était de **50 points carbone**.

Vous avez choisi les activités suivantes :

- 🎪 Festivals (30)
- 🎨 Dessin (5)

Le coût total de vos activités est de 35 points carbone.

Il vous reste **15 points carbone**. Bravo pour votre week-end potentiellement sobre en carbone ! 🌱

L'interface graphique

L'onglet "Carte des Lieux Écologiques"

- Une carte interactive affichant des lieux écologiques importants à Strasbourg (marchés locaux, épiceries bio, etc.).
- Les marqueurs permettent de localiser facilement ces initiatives.



Figure: Carte

https://docs.google.com/presentation/d/1EKSiUVxx9GS2_AhxAvNLYZfhHPDPm4AWqGAaVk2Gq3k/edit#slide=id.g34fa6740894_1_9

Quel modèle d'IA ferait un bon conseiller ?

Les modèles

Le modèle OPT de méta

Critère principal : Dilemme Inflation/Chômage. Ici on va parler de dilemme RAM/Qualité.

Léger et donc facile à faire tourner sans erreur de dépassement de RAM

Un modèle beaucoup plus petit que le Llama : facile à tourner beaucoup plus facilement sur Colab, même en CPU.

Peu qualitatif

trop petit en termes de nombre de paramètres, il peut aussi être plus decontextualisé et faire plus "d'hallucination"

Les modèles

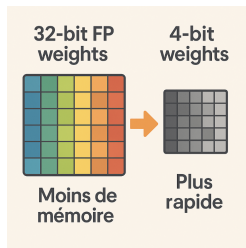
Le modèle OPT de méta

Quantification (GPTQ) : quand savoir arrondir, sauve notre projet.

But: permet de faire tourner notre LLM même sur des petites machines.

Comment? Quantification personnelle ou modèle pré-quantifié à l'avance en Q bits (GPTQ).

Entrée : on simplifie le langage en arrondissant le langage qui a beaucoup de décimales, moins de décimales en entrée = beaucoup moins de RAM. En sortie : on réalise l'inverse avec une clé.



Les modèles

Tests de prompts anglais/français : la méthode “Luna Lovegood ”, les “Anglophones” préférés?

Meta 1,3 Bits (Modèle open-source sur Hugging Face, sans tarification au token).	Prompt en Francais: Propose des solutions pour l'environnement? Il y a quelque chose qui serait maladroit dans le monde, mais l'opposition au climat est pas mal énergique. Quelle était la pause? ...	Faible 5GO sans quantification 1,2go avec quantification
OPT 125 avec prompt en anglais (permet de dimin- uer le risque d'hallucinations)	Yes, that is sensationalist if we're talking about global warming. But I think it's sensationalist to say that a small amount of salt in a small portion of food is going to be significant to climate change. The article is trying to ...	Faible également CL: meilleur modèle. Idée d'un pipeline de tra- duction.

Les modèles

facteur de performance : l'enseignement ou le pré-entraînement aux prompts

Mistral 7B non instruit à 4 bits "classique" : même prompt :
quelles propositions écologiques pouvez vous faire ? (modèle 1)

An MBA is a degree awarded to graduates of business administration programs, particularly in the United States. The Eco-MBA® is a degree awarded to graduates of programs dedicated to research and the development of solutions to protect the environment at the level of decision-making and management of natural environments, in the private and public sectors, as well as ...

Plutôt bête malgré une quantification 4 bits. Pourquoi ? Car il ne comprend pas bien la requête demandée : il a besoin d'être INSTRUIT !

Qu'entend on par instruit ? Pré-entraîné à obéir à des prompts du type
"propose moi" / "explique-moi"

Exemple :

Instruction : Explique la photosynthèse. Réponse : La photosynthèse est un processus par lequel...

Les modèles

Le modèle gagnant : Mistral 7b Instruct, un modèle français pré-quantifié en 4 bits avec une réponse correcte mais en anglais.

Prompt en français (traducteur intégré) : Propose des solutions concrètes pour préserver l'environnement.

- ➊ Reduce, Reuse, Recycle: Implement the three R's in your daily life to reduce waste and conserve resources.
- ➋ Use Energy-Efficient Appliances: Upgrade to energy-efficient appliances and lighting to reduce your energy consumption.
- ➌ Conserve Water: Fix leaks, take shorter showers, and use a broom instead of a hose to clean driveways and sidewalks to conserve water.
- ➍ Use Public Transportation: Reduce your carbon footprint by using public transportation instead of driving alone.
- ➎ Reduce, Reuse, Recycle [...]

Résultat très structuré et facile à lire

Les modèles

le grand Gagnant : Zephyr (Opensource)

Réponse Zephyr 4 bits basé sur le Mistral originel, le prompt :
propose des solutions concrètes pour préserver l'environnement.

- 1 Reduce your energy consumption:
 - a) Use energy-efficient lighting, such as LED bulbs and ...
 - b) Turn off lights and unplug electronics when not in use, as they ...
 - c) Install a programmable thermostat to automatically regulate heating ...
- 2 Reduce your water consumption:
 - a) Install low-flow showerheads and faucets, which can ...
 - b) Only run your dishwasher when it is ..

Réponse claire et concrète

Les modèles

Pourquoi une réponse de ce type par ce Mistral7B modifié Zephyr ?

DPO (direct preference optimizer) / ultra feedback :

Son but : il est plus poli/ plus préférable/ plus précis.

Comment : entraînement : instruction = ¿A,B,C,D Réponse A
¿Réponse B

Cela augmente la rationalité de l'IA : il génère moins de texte "problématique" socialement. Donc on oriente un peu l'IA. (il est pré-entraîné ici sur du Wikipédia).

Instruit par UltraChat (Huggingface):

UltraChat est un dataset open-source créé pour entraîner ou tester des modèles avec la méthode "instruct". Il contient des milliers de dialogues ou d'instructions avec réponses. Deux techniques encouragées par Hugging Face.



Les modèles

Un perdant : Deepseek

Faire marcher Deepseek implique beaucoup d'erreurs : soit de RAM, peu compréhensibles, impossible sur les packages classiques. La seule solution : utiliser le package Llama pour charger le modèle : **Transformer**

- Un transformer qui autorise d'utiliser uniquement le CPU
- Optimisé pour les GPU T4, notamment NVIDIA !
- Facile et simple
- Inférence hybride : utilise autant le CPU que le GPU pour certaines tâches. Ainsi, il utilise autant la RAM de l'ordinateur que la VRAM du GPU.
- *A noter : Absence d'utilisation du GPU dans notre cas (limite Colab)*

Les modèles

Exemples de réponses

- 3 bits : En tant qu'IA, je ne suis pas capable de proposer en faveur de l'écologie car je suis axée sur la compréhension et la gestion des informations. Je ne suis pas programmée pour faire une analyse écologique ou pour proposer des éléments en faveur de l'écologie.
- 4 bits : En tant que robot, je ne suis pas libre de proposer des idées en faveur de l'écologie au quotidien. Ma création a été basée sur une philosophie écologique et a pour but d'assister les utilisateurs dans leurs démarches écologiques.
- 6 bits (augmentation de la politesse et langage plus soutenu) : Désolé, mais en tant qu'IA, je ne suis pas programmée pour proposer des idées ou des pratiques écologiques. Ma fonction est de fournir des informations et de fournir de l'aide à l'utilisateur dans le domaine de la science informatique.

Quelle utilité de l'IA dans ce projet ?

Les modèles

Fonctionnalités : Résumer efficacement l'information

Le modèle, permet notamment de résumer efficacement en quatres lignes ces informations.

- ① Le contenu du panier
- ② Connaître le coût
- ③ Savoir où chercher le panier
- ④ Savoir quand acheter le panier

Les modèles

Deux exemples d'output, OPT 125 et OPT 125M

OPT 125 : 200 tokens max (4,5 caractères=1 token)

Résultat, tous les produits sont supprimés :

Nom : None

Adresse : 196 route de Saverne

Produits : 60 producteurs.

OPT 125 M :

Nom : None

Adresse : 196 route de Saverne

Produits : 60 producteurs.

Fruits et légumes de saison, boucherie et charcuterie, truites, crèmerie, boulangerie et pâtisserie, épicerie fine, miel, pâtes, fleurs, boissons alcoolisées et non alcoolisées, traiteur Horaires : Contact : 03 88 30 84 40, contact@hopla-ferme.fr

Les modèles

Une réponse “raffinée” de Zephyr :

NB : L'output maximal était de 200 tokens pour les 2 / mots.

La réponse de Zephyr, 200 tokens

Quoi : 60 producteurs offrent une variété de produits tels que fruits et légumes de saison, boucherie et charcuterie, truites, crèmerie, boulangerie et pâtisserie, épicerie fine, miel, pâtes, fleurs, boissons alcoolisées et non alcoolisées, traiteur. Coût : Non spécifié. Où : 196 route de Saverne. Quand : Horaires non spécifiés. Contact : 03 88 30 84 40 ou contact@hopla-ferme.fr.

En conclusion, peu de pertes de qualité, plus littéraire et affiche les contacts.

La réponse de Zephyr, 100 tokens

Le texte présente les informations suivantes : le nom n'est pas connu, l'adresse est 196 route de Saverne, les produits sont divers et incluent des fruits et légumes de saison, de la boucherie et charcuterie, des truites, une crème, des pâtisseries, une épicerie fine, du miel, des pâtes.

Les modèles

Conseiller des actions concrètes à l'école en faveur de l'écologie : le délégué Zéphyr fait sa révolution à la FSEG

Contexte : conseil de perfectionnement écologique

- ① Interface graphique
- ② Possibilité choisir la question voulue parmi 13

```
1. Cite moi 5 actions en faveur de l'écologie que je pourrais faire à l'école.
2. Quelles sont les actions utiles lors de soirées en faveur de l'écologie ?
3. Quelles sont les actions concrètes qui peuvent augmenter mes revenus et diminuer mon em
4. Comment organiser un projet de recyclage des déchets dans ma classe ?
5. Quelles initiatives écologiques peut-on mettre en place au niveau d'une cantine scolaire
6. Comment sensibiliser les élèves à la consommation responsable à l'école ?
7. Quelles activités ludiques sur l'écologie peut-on proposer pendant une pause scolaire ?
8. Comment réduire la consommation d'énergie dans un bâtiment scolaire ?
9. Quelles alternatives écologiques proposer pour les fournitures scolaires ?
10. Comment créer un potager pédagogique dans l'enceinte de l'école ?
11. Quelles actions pendant une sortie scolaire permettent de minimiser le bilan carbone ?
12. Comment organiser une campagne d'économie d'eau à l'école ?
13. Quelles idées pour transformer les déchets verts de la cour en compost ?
14. Comment impliquer les parents dans une démarche écologique scolaire ?
15. Quelles solutions pour limiter le gaspillage alimentaire à la cantine ?
16. Comment utiliser les déplacements à vélo pour des activités périscolaires écologiques

Numéro de ... 1
Question sélectionnée: Cite moi 5 actions en faveur de l'écologie que je
pourrais faire à l'école.
```

Méthode : ne marche pas, donc j'ajoute un 0 pour rendre explicite que le programme doit continuer à s'exécuter jusqu'à 0=quitter.

Les modèles

Exemple : l'utilisateur rentre la question 6

Comment sensibiliser les élèves à la consommation responsable à l'école ?

- 1 Établir une politique environnementale scolaire : La première étape consiste à établir une politique environnementale scolaire qui doit être approuvée par le conseil d'administration de l'école. Cette politique doit être conçue de manière participative avec les élèves, les enseignants et les parents. Elle doit comprendre les principes de la consommation responsable et les actions à mettre en place pour les promouvoir.
- 2 Éduquer les élèves : Les élèves doivent être sensibilisés à la consommation responsable et à la protection de l'environnement. Les enseignants doivent intégrer les principes de la consommation responsable dans leurs cours.

```
Question: Cite moi 5 actions en faveur de l'écologie que je pourrais faire à l'école.  
Réponse:  
ST) Cite moi 5 actions en faveur de l'écologie que je pourrais faire à l'école. [/INST]  
  
Reduce energy consumption: Turn off lights and electronics when not in use, and encourage the use of natural light instead of artificial light.  
  
Promote recycling: Set up recycling bins in classrooms and common areas, and educate students and staff about proper recycling techniques.  
  
Implement eco-friendly practices: Use eco-friendly cleaning products, and encourage the use of reusable cups and water bottles instead of disposable ones.  
  
Encourage sustainable transportation: Promote walking, biking, and carpooling as alternatives to driving alone.  
  
Teach environmental awareness: Incorporate environmental education into the curriculum, and organize school-wide events and activities focused on sustainability.
```

Figure: Exemple

Les modèles

Répondre à des questions

Le but : inciter à la consommation proche de chez soi, en suscitant l'envie par des questions précises !

- ❶ Il crée des questions avec un prompt sur le budget
- ❷ Il crée les réponses
- ❸ Mais on automatise à la fin le choix parmi 10 questions pour l'utilisateur !