

Fiabilité des données de l'IoT + Classification des intentions basée sur les LLM et optimisation du réseau

Présentation MESPI

Par:

AIT HAMOUDA Gaya

CHARFI Sirine

ISSAAD Célia

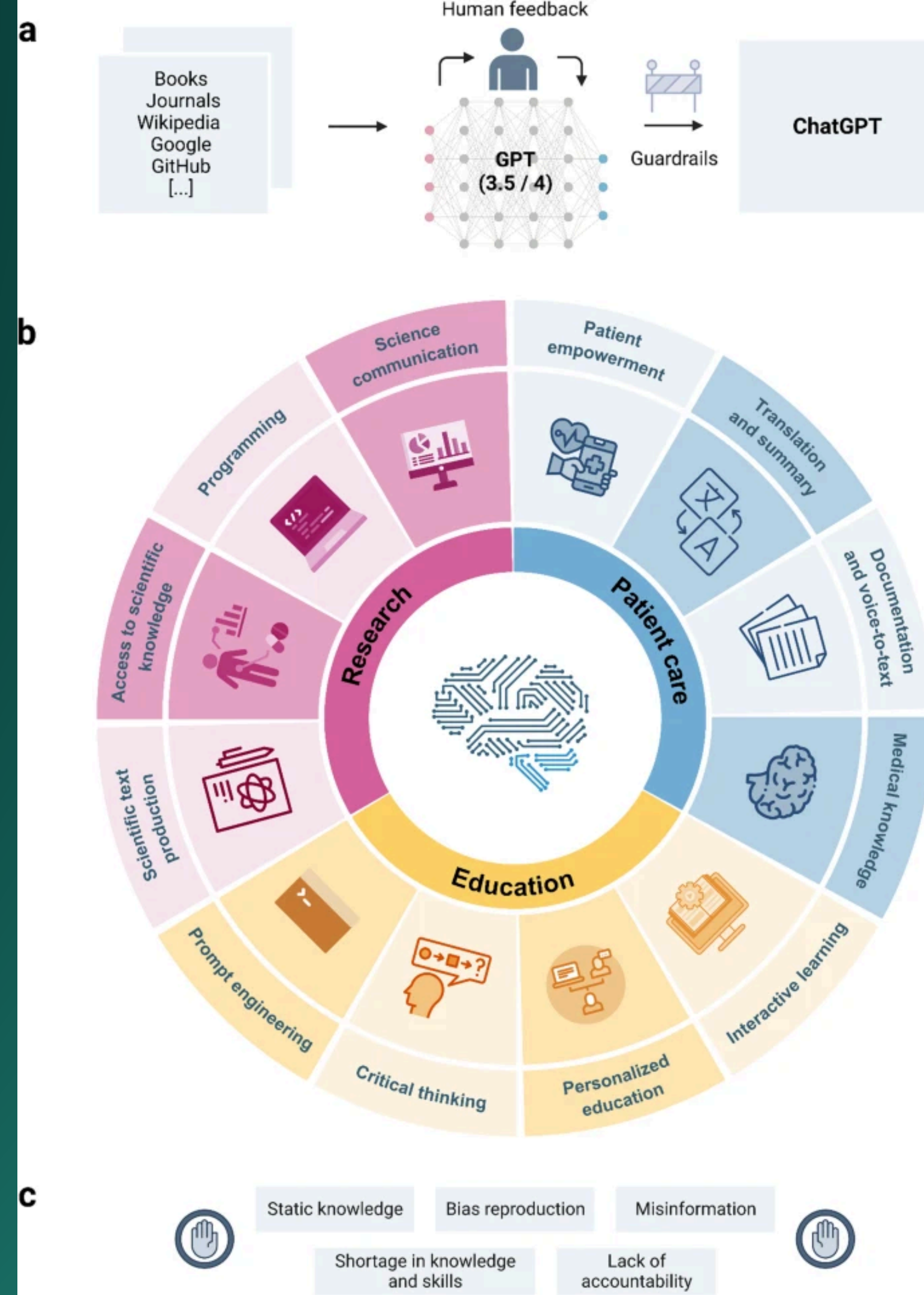
Introduction

Objectif :

- Améliorer la prise de décision dans la classification des données médicales et la structuration des intentions
- Utilisation de modèles de traitement du langage naturel (Mistral, LLaMA, DeepSeek v3) pour classer les intentions des utilisateurs et optimiser les ressources réseau.

Pourquoi c'est important :

- Des données fiables sont essentielles dans le domaine de la santé
- Comprendre les intentions des utilisateurs améliore l'automatisation et la qualité de service (QoS) des réseaux



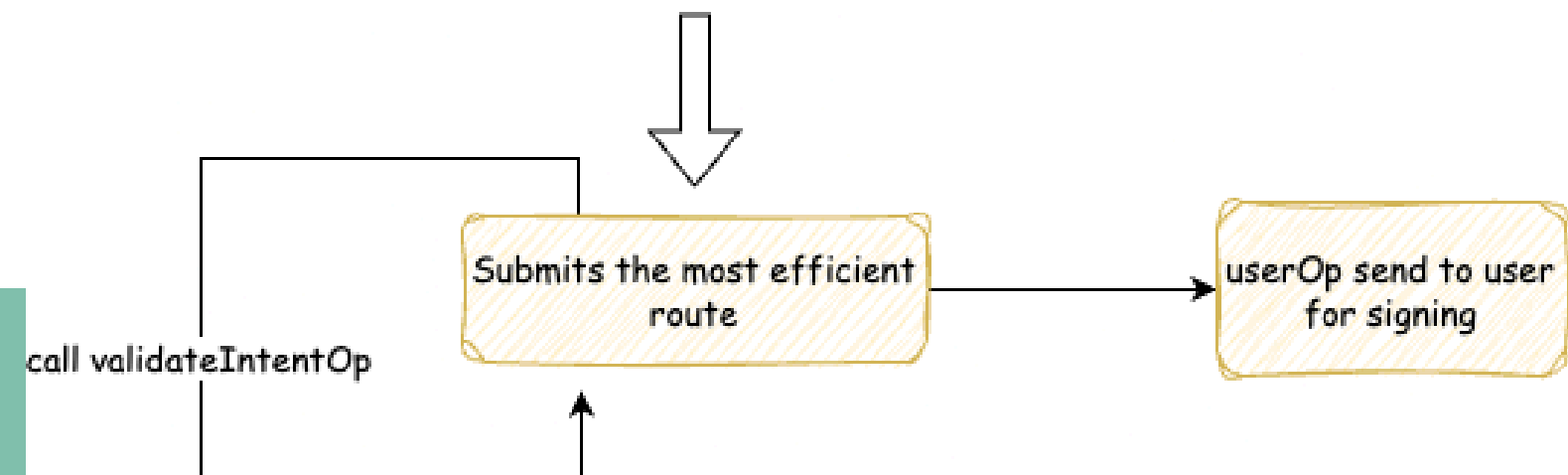
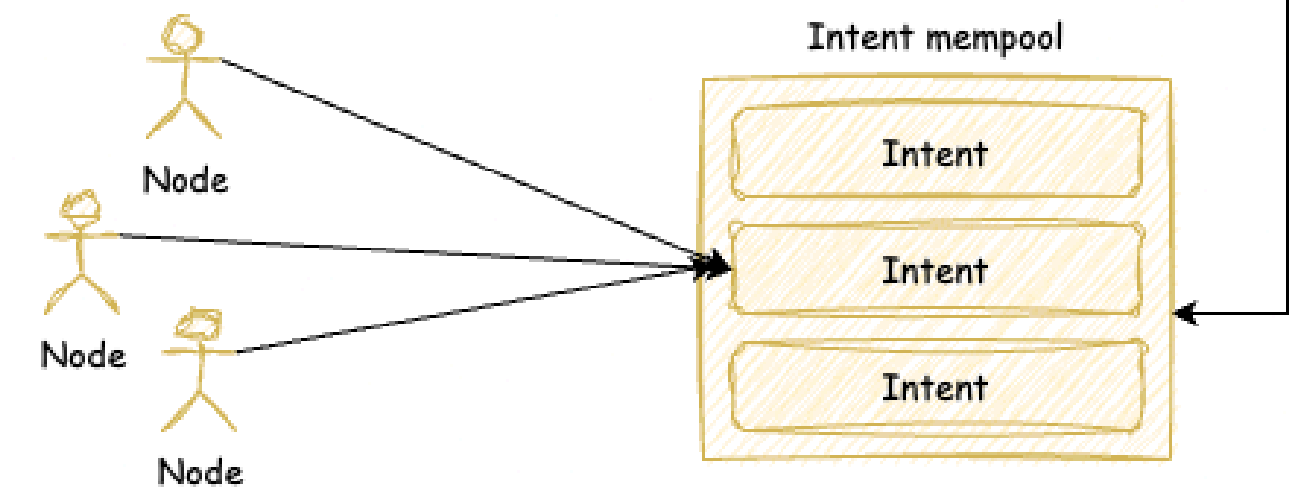
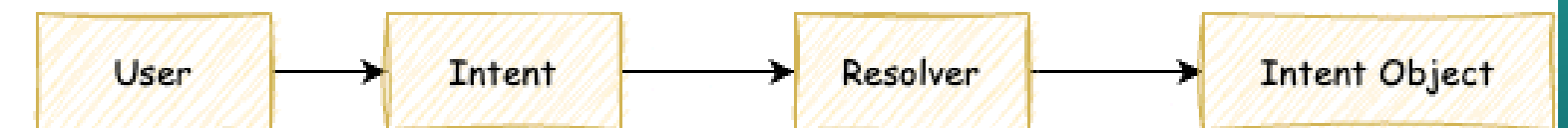
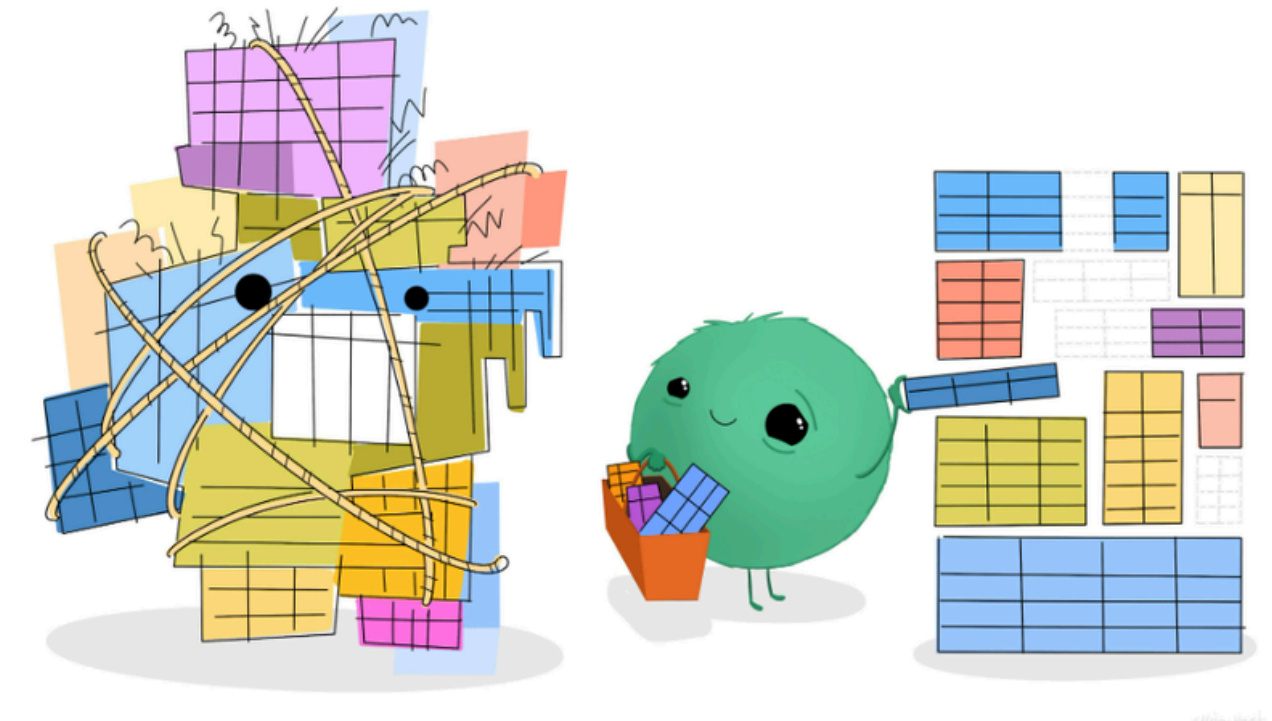
Contexte et énoncé du problème

Défis liés aux données médicales :

- Sources de données hétérogènes et peu fiables
- Risques de classifications incorrectes (mauvaises prescriptions, diagnostics erronés).

Structuration des intentions dans les réseaux :

- Les utilisateurs formulent des requêtes en langage naturel.
- Les systèmes doivent les comprendre, les classer et optimiser les réponses de manière dynamique.



Modèles LLM pour la classification des intentions :

- Convertir le langage naturel en intentions structurées.
- Optimiser les ressources réseau en fonction des requêtes des utilisateurs.

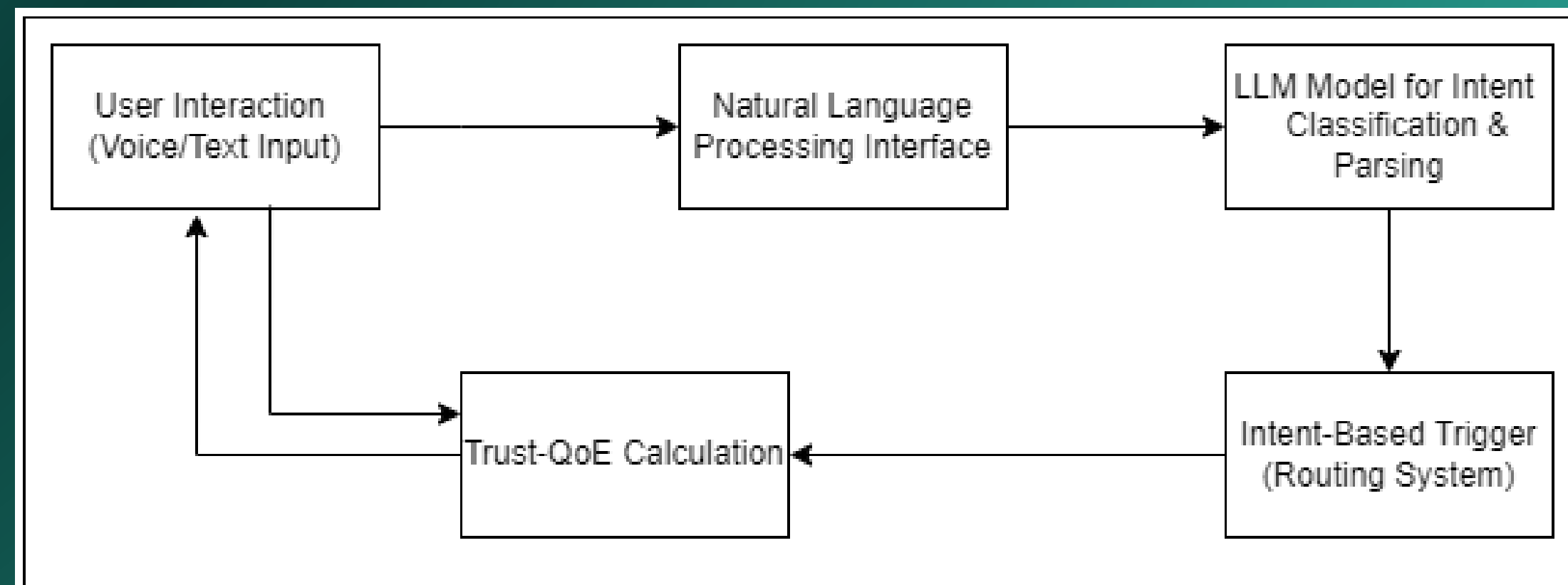
Implémentation du modèle LLM dans une application :

- Utilisation des résultats du modèle LLM pour surveiller les activités réseau (et d'autres cas d'utilisation si possible)

Objectif final :

Améliorer la prise de décision dans le domaine de la santé et l'automatisation des réseaux

Objectifs du projet



Methodology

1. Création du Dataset

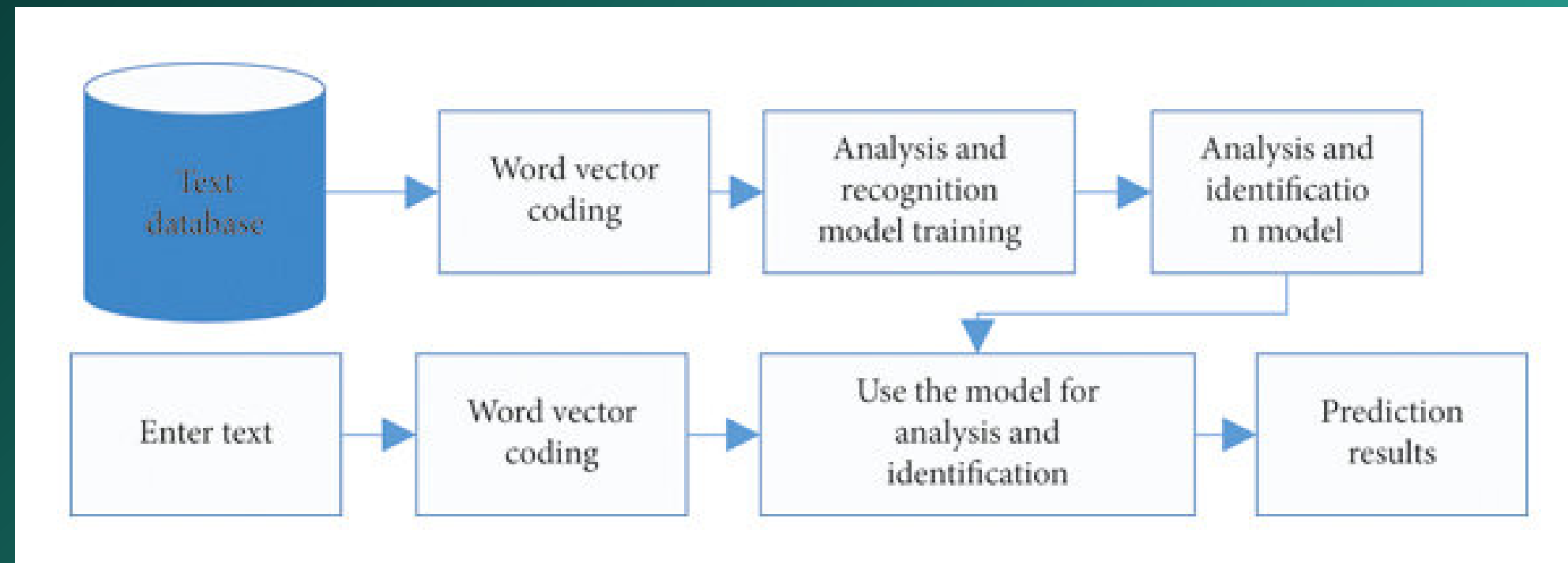
- ~160 exemples d'entrées utilisateur + intentions structurées

Ex :

```
{
  "input": "I want to transfer patient vitals to the cloud securely.",
  "intent": {
    "action": "ensure connectivity",
    "category": "healthcare data transfer",
    "goal": "secure connection",
    "service": "healthcare cloud storage",
    "service_type": "GDPR-compliant service",
    "network": "dedicated healthcare network",
    "device": "patient monitoring device",
    "device_capability": "supports secure communication",
    "context": "hospital environment",
    "time_sensitivity": "moderate",
    "user_profile": "nurse",
    "data_preference": "ensure security",
    "priority": "secure connection",
    "energy_saving": true,
    "security_level": "GDPR-compliant",
    "QoS_attribute": {
      "bandwidth": "high",
      "latency": "low"
    },
    "expected_metrics": {
      "bandwidth": "at least 10 Mbps",
      "latency": "<50ms"
    }
  }
}
```

2. Différentes approches

- **Zero-shot learning** : Modèle brut, sans entraînement.
- **Few-shot learning** : Ajout d'exemples.
- **Fine-tuning** : Entraîné sur données personnalisées.
- **Retrieval-Augmented Generation (RAG)** : Résultats améliorés avec données externes.



Results

Modèles envisagés :

- Mistral 📊 (précision modérée attendue, réponse rapide)
- LLaMA 📊 (performance équilibrée attendue)
- DeepSeek v3 📈 (meilleure précision attendue)

Modèles testés :

- LLaMA-2-7B
- Mistral-7B
- Mistral-7B-v0.1
- LLaMA-2-13B
- LLaMA-2-70B (impossible sur Colab)
- Deepseek-70B (impossible sur Colab)

Problèmes :

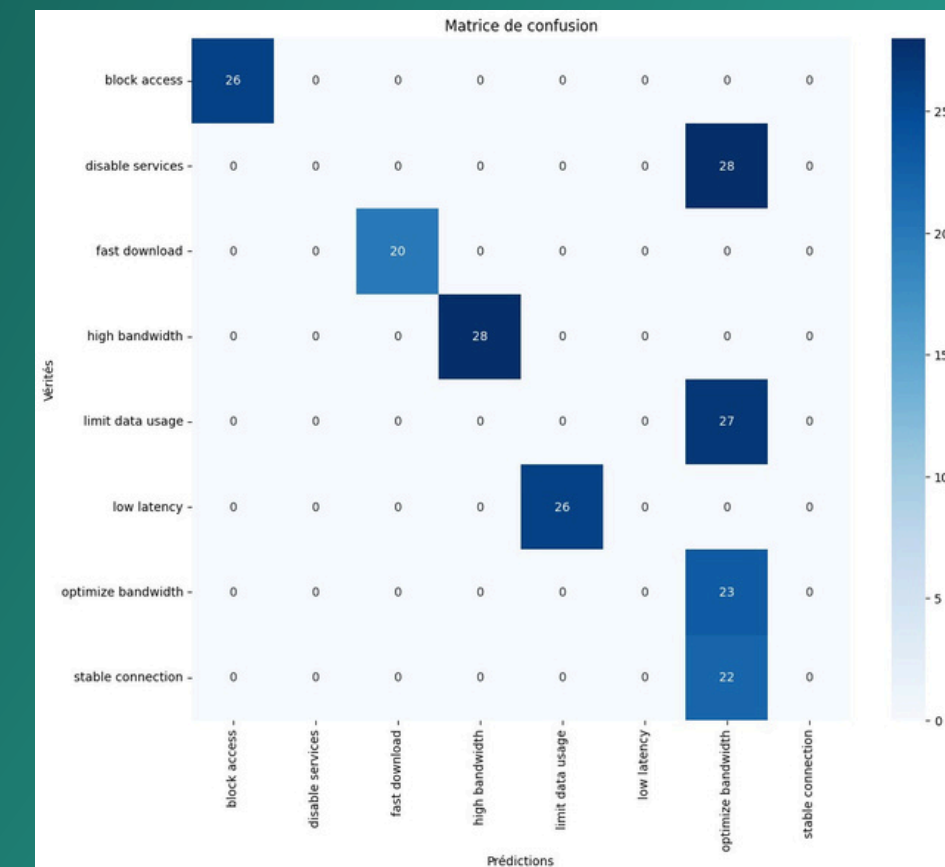
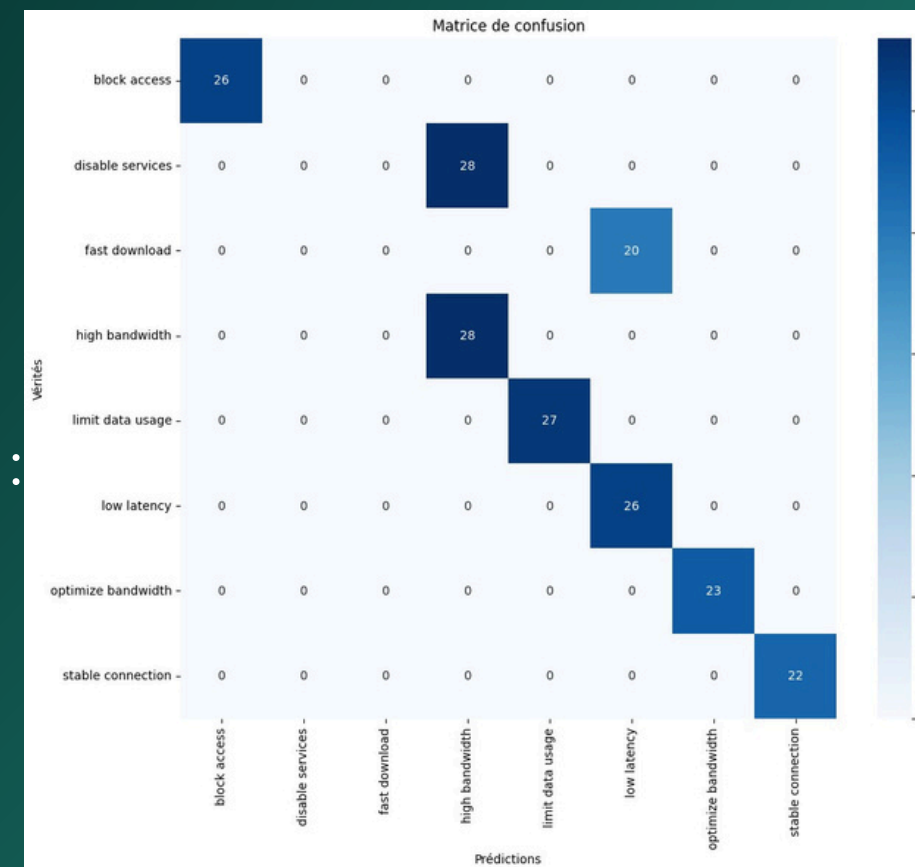
- Plusieurs modèles testés, mais toujours des soucis, dont :
- RAM insuffisante sur GPU T4 de Google Colab
- Les modèles fonctionnant sur Colab sont trop faibles pour de bons résultats

Solution :

- Ordinateur avec bon GPU (RTX A5000) fourni pour tests, en cours d'utilisation.
- Environnement ollama pour tests.

Solution alternative :

- Si échec avec modèles open source, tentative avec GPT4 pour la tâche de classification d'intentions (utilisation d'une API OpenAI pour intégrer GPT-4 à notre application)



Résultats

Résultats de GPT4 :

- 100 % de précision pour toutes les fonctionnalités principales des entrées de notre jeu de données (action, objectif, catégorie, service).

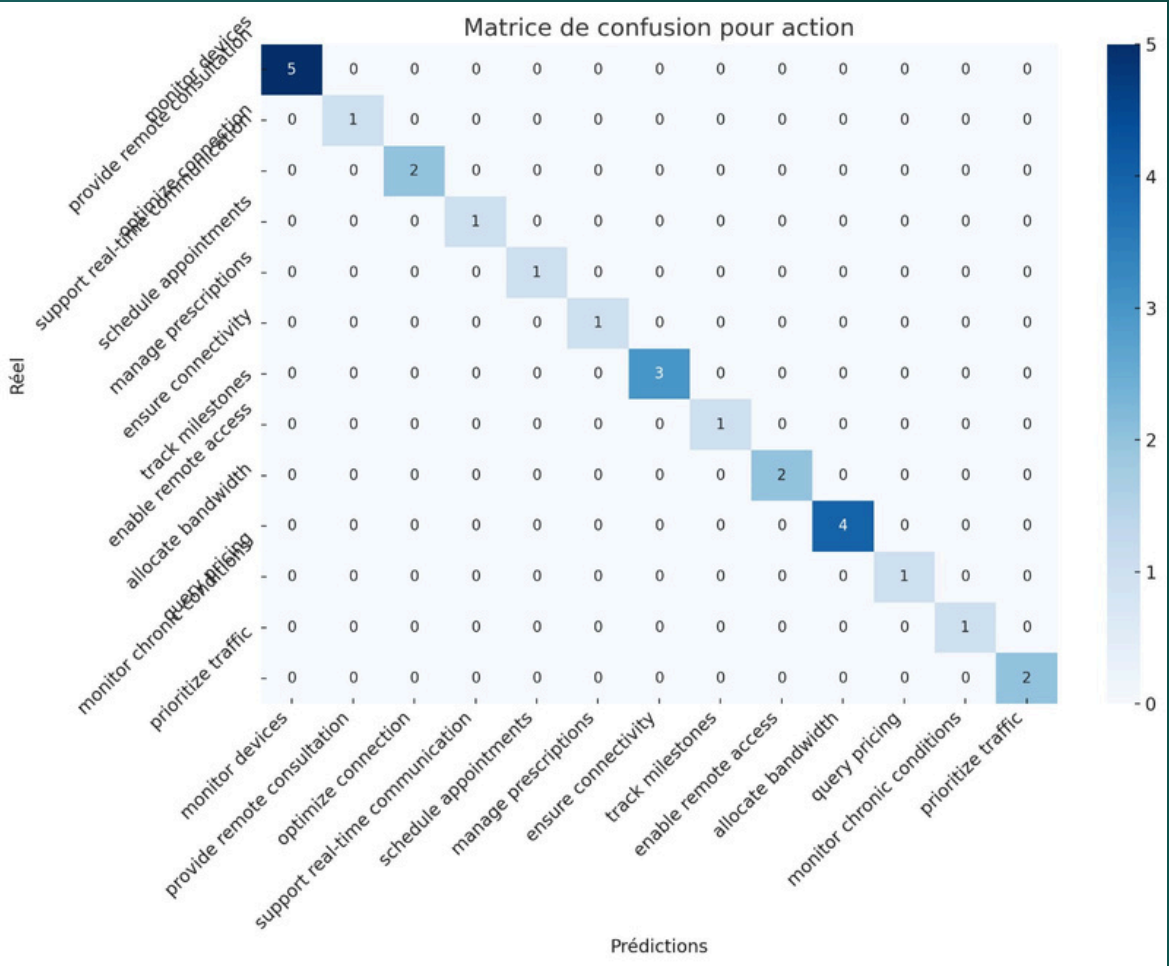
Conclusion :

- Indique clairement que le prompt n'était pas le problème (comme suggéré), mais que les modèles utilisés jusqu'à présent n'étaient pas assez puissants.

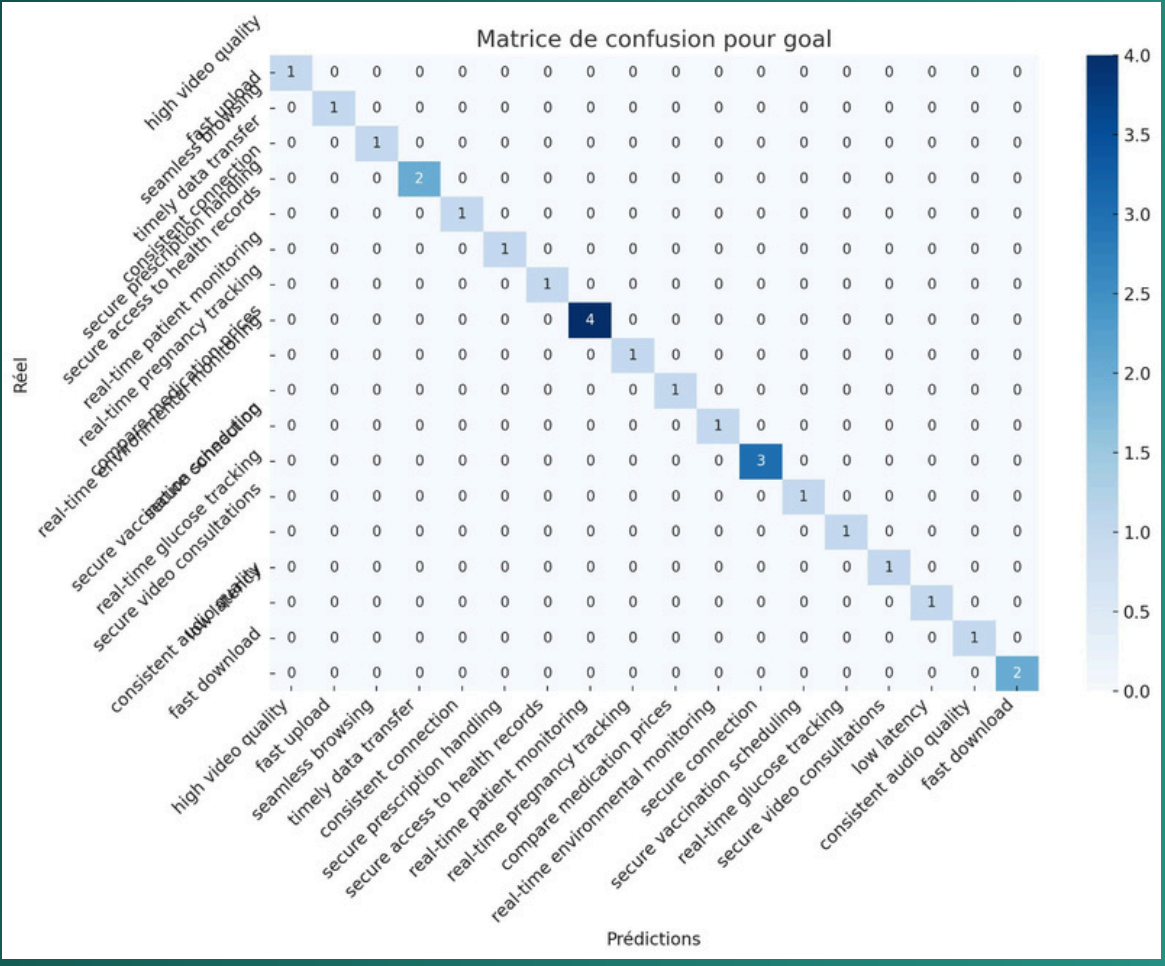
Idée :

- Utiliser ChatGPT via l'API OpenAI pour créer notre propre application et prototype pour l'optimisation de réseau (entre autres cas d'usage médicaux).

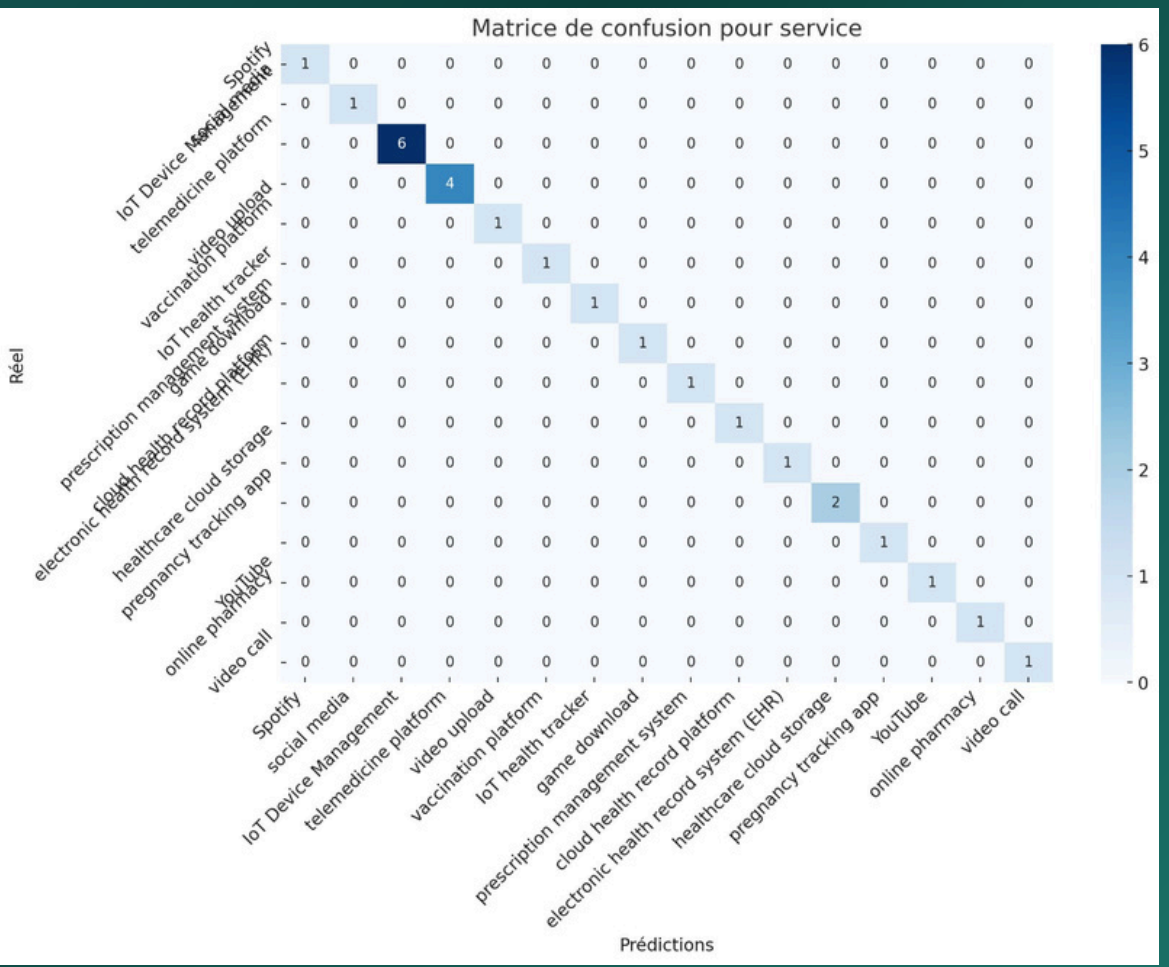
Action



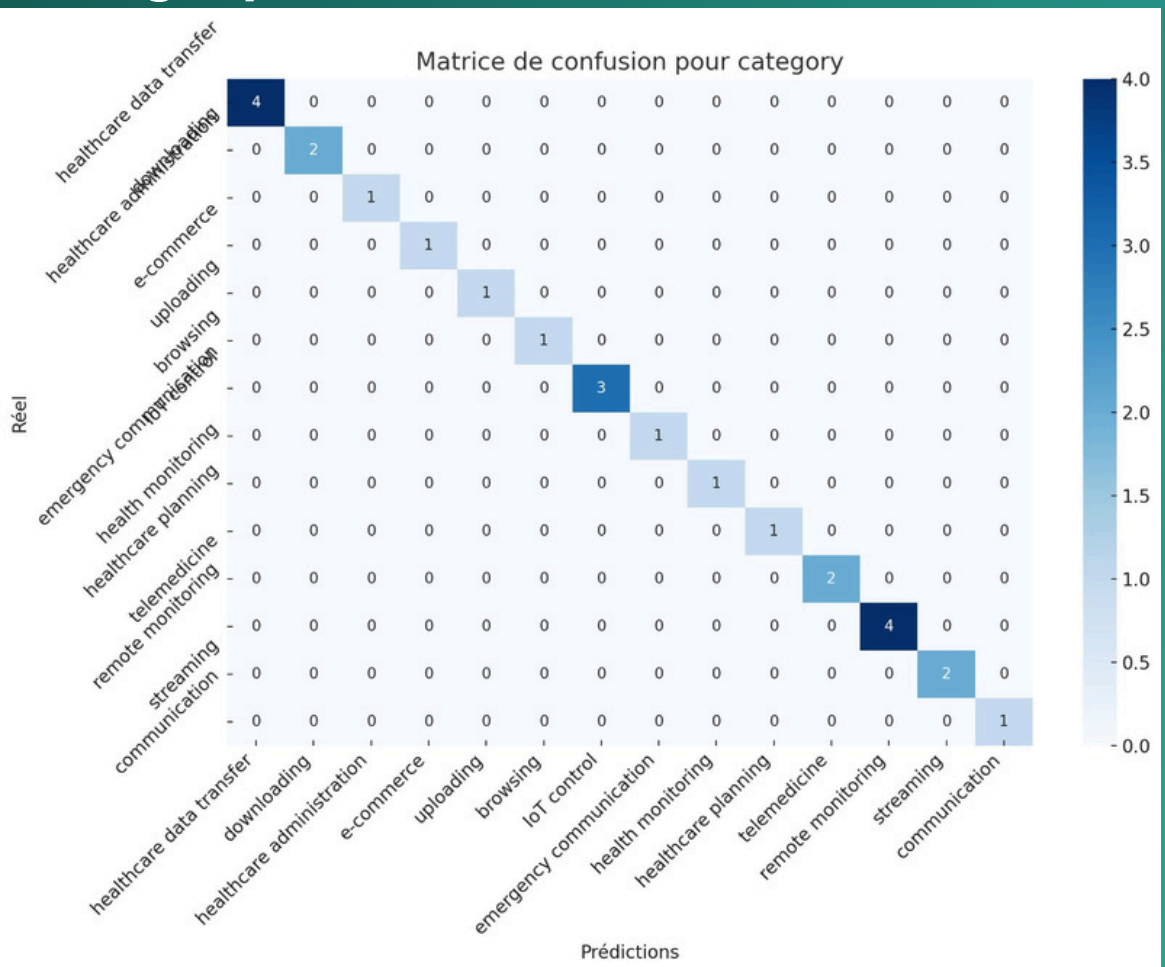
Goal



Service



Category



Merci