

Développement d'un Chatbot Intelligent Spécialisé sur les Politiques Publiques Sénégalaises

Abdoulaye SALL

Résumé

Ce rapport présente le développement d'un chatbot intelligent spécialisé dans les politiques publiques sénégalaises.

Utilisant l'architecture RAG (Retrieval-Augmented Generation), ce système offre un accès facile et direct aux informations concernant l'Agenda National de Transformation Sénégal 2050, la Stratégie Nationale de Développement 2025-2029, et le New Deal Technologique.

Le projet s'inscrit dans une démarche de démocratisation de l'accès à l'information gouvernementale, permettant aux citoyens, chercheurs, étudiants et professionnels de mieux comprendre les orientations stratégiques du Sénégal.

Table des matières

1	Introduction	4
1.1	Contexte du projet	4
1.2	Objectifs du projet	4
2	Vue d'ensemble du système	5
2.1	Principe de fonctionnement	5
2.2	Composants principaux	5
3	Ingestion et préparation des données	6
3.1	Sources documentaires	6
3.2	Processus de traitement	6
4	Système de récupération d'information	7
4.1	Base vectorielle	7
4.2	Mécanisme de récupération	7
5	Intelligence artificielle et génération de réponses	8
5.1	Configuration du modèle	8
6	Interfaces utilisateur et API	9
6.1	Application Streamlit	9
6.2	Interface API	9
7	Impact et perspectives	10
7.1	Impact sociétal	10
7.2	Perspectives d'évolution	10
8	Conclusion	11

1 Introduction

1.1 Contexte du projet

Le Sénégal a récemment connu un changement majeur de gouvernance avec l'arrivée à la présidence de Son Excellence Bassirou Diomaye Faye et la nomination d'Ousmane Sonko au poste de Premier Ministre.

Cette nouvelle administration a introduit des plans stratégiques ambitieux pour transformer le pays, notamment :

- L'Agenda National de Transformation Sénégal 2050
- La Stratégie Nationale de Développement 2025-2029
- Le New Deal Technologique

Ces documents définissent une vision claire pour l'avenir du Sénégal, mais leur contenu peut souvent être difficile d'accès pour le grand public.

Les informations sont dispersées dans de longs documents officiels, complexes et parfois techniques pour certains, ce qui limite la compréhension et l'appropriation des politiques par les citoyens.

1.2 Objectifs du projet

Face à ce constat, ce projet vise à :

- Démocratiser l'accès aux informations gouvernementales
- Offrir un outil interactif qui répond aux questions spécifiques des utilisateurs
- Fournir des réponses précises et documentées issues des sources officielles
- Promouvoir la transparence des politiques publiques
- Faciliter la compréhension des orientations stratégiques du pays

Le chatbot développé se positionne comme un médiateur entre les politiques gouvernementales et les citoyens, contribuant ainsi à une meilleure gouvernance participative et démocratique.

2 Vue d'ensemble du système

2.1 Principe de fonctionnement

Le chatbot repose sur le paradigme RAG (Retrieval-Augmented Generation), qui combine deux approches :

Architecture RAG

1. **Retrieval (Récupération)** : Le système recherche les informations pertinentes dans une base documentaire préalablement indexée.
2. **Augmented Generation (Génération Augmentée)** : Un modèle de langage (LLM) génère des réponses précises en se basant sur les informations récupérées.

Cette approche garantit des réponses à la fois précises (car basées sur des documents officiels) et naturelles (grâce aux capacités de formulation du LLM).

2.2 Composants principaux

Le système s'articule autour de quatre composants majeurs, chacun jouant un rôle essentiel dans la chaîne de traitement :

1. **Module d'ingestion** : Processus de chargement et préparation des documents officiels
2. **Base de connaissances vectorielle** : Stockage optimisé pour la recherche sémantique
3. **Moteur de récupération** : Identification des passages pertinents pour chaque question
4. **Interface utilisateur** : Point d'interaction directe avec les utilisateurs

3 Ingestion et préparation des données

3.1 Sources documentaires

Le système exploite trois documents stratégiques fondamentaux :

- **Agenda Sénégal 2050** : Vision à long terme du développement national
- **Stratégie Nationale 2025-2029** : Plan quinquennal de développement
- **New Deal Technologique** : Stratégie de transformation numérique

Ces documents constituent le socle informationnel du chatbot, garantissant des réponses basées sur des sources officielles et faisant autorité.

3.2 Processus de traitement

L'ingestion des documents suit un processus rigoureux en plusieurs étapes :

1. **Extraction de texte :**

Convertit les documents PDF en texte brut exploitable, en préservant au maximum la structure originale.

2. **Prétraitement textuel :**

Nettoie et normalise le texte pour améliorer la qualité de l'indexation (suppression des caractères spéciaux, normalisation des espaces).

3. **Segmentation contextuelle :**

Découpe intelligemment les documents en segments cohérents, préservant le contexte sémantique des informations.

4. **Indexation vectorielle :**

Convertit les segments textuels en représentations vectorielles et les stocke dans une base optimisée pour la recherche sémantique.

Ce processus transforme des documents bruts en une base de connaissances structurée et interrogeable, fondement de la pertinence des réponses du système.

4 Système de récupération d'information

4.1 Base vectorielle

La base vectorielle ChromaDB constitue le cœur du système de récupération :

Avantages de ChromaDB

- Stockage optimisé des représentations vectorielles des segments de texte
- Recherche par similarité sémantique ultrarapide
- Persistance des données pour un accès instantané
- Évolutivité pour intégrer de nouveaux documents

Cette technologie permet de dépasser les limitations des recherches par mots-clés traditionnelles en identifiant les passages sémantiquement proches de la question, même lorsque les termes exacts diffèrent.

4.2 Mécanisme de récupération

Le module de récupération joue un rôle critique :

- Analyse la question de l'utilisateur pour en extraire l'intention
- Convertit cette question en requête vectorielle
- Interroge la base ChromaDB pour identifier les segments les plus pertinents
- Sélectionne un nombre optimal de passages pour constituer le contexte de réponse

La qualité de cette récupération conditionne directement la précision des réponses générées.

Des mécanismes de paramétrage permettent d'ajuster la sensibilité de la recherche selon les besoins.

5 Intelligence artificielle et génération de réponses

5.1 Configuration du modèle

Le système s'appuie sur Gemini 1.5 Flash de Google pour la génération de réponses :

Caractéristiques du modèle

- LLM de dernière génération avec compréhension contextuelle avancée
- Capacités multilingues (français, anglais)
- Formulation naturelle et adaptée au contexte
- Optimisation temps/qualité pour des réponses rapides mais précises

L'ingénierie de prompt joue un rôle crucial dans l'optimisation des réponses. Les instructions fournies au modèle sont soigneusement calibrées pour :

- Cadrer le contexte (politiques publiques sénégalaises)
- Définir le ton (informatif, objectif)
- Préciser les limites (se baser uniquement sur les documents fournis)
- Structurer la réponse (clarté, concision)

6 Interfaces utilisateur et API

6.1 Application Streamlit

L'interface développée avec Streamlit offre une expérience intuitive et accessible :

Caractéristiques de l'interface

- Design épuré et convivial adapté au contexte
- Organisation thématique des questions (Souveraineté Économique, Technologie Numérique, Justice Sociale)
- Vérification de la pertinence thématique des questions
- Affichage clair et structuré des réponses

Cette interface web est conçue pour être accessible au plus grand nombre, sans nécessiter de compétences techniques particulières.

6.2 Interface API

En complément de l'interface web, une API (fastapi) permet l'intégration du chatbot dans d'autres applications :

- Endpoint unique pour soumettre des questions
- Format JSON standardisé pour les requêtes/réponses
- Documentation interactive via Swagger UI
- Gestion robuste des erreurs

Cette API ouvre la voie à des intégrations diverses, applications web et mobiles.

7 Impact et perspectives

7.1 Impact sociétal

Le chatbot représente bien plus qu'un simple outil technologique ; c'est un vecteur de transformation sociale :

Impacts attendus

- **Démocratisation de l'information** : Accès facilité aux politiques publiques pour tous les citoyens
- **Transparence gouvernementale** : Clarification des orientations stratégiques nationales
- **Éducation civique** : Meilleure compréhension des enjeux politiques et sociétaux
- **Participation citoyenne** : Encouragement au débat public informé
- **Réduction des inégalités informationnelles** : Accès aux mêmes informations pour tous

En rendant accessibles les informations gouvernementales, ce chatbot contribue au renforcement de la démocratie sénégalaise et à l'implication citoyenne dans les affaires publiques.

7.2 Perspectives d'évolution

Le projet ouvre de nombreuses pistes d'amélioration et d'extension :

1. **Multilinguisme** : Extension aux langues nationales (Wolof, Pulaar, Sérère)
2. **Visualisation** : Intégration de graphiques pour illustrer les indicateurs clés
3. **Veille automatique** : Mise à jour continue avec les nouvelles publications officielles
4. **Version mobile** : Application dédiée pour smartphones et tablettes
5. **Personnalisation** : Adaptation des réponses selon le profil de l'utilisateur
6. **Intégration multimédia** : Incorporation de vidéos et discours officiels

Ces évolutions permettraient d'accroître encore l'impact et l'accessibilité du système.

8 Conclusion

Le développement d'un chatbot spécialisé sur les politiques publiques sénégalaises peut être nécessaire dans le paysage de l'information gouvernementale au Sénégal.

En combinant des technologies avancées d'intelligence artificielle (RAG, LLM) avec une connaissance approfondie du contexte local, ce projet propose une solution concrète à un défi majeur : rendre compréhensibles et accessibles des orientations stratégiques complexes mais essentielles pour l'avenir du pays.

Au-delà de sa dimension technologique, ce chatbot incarne une vision de la gouvernance où l'information n'est plus un privilège mais un droit accessible à tous.

Il contribue ainsi à construire un Sénégal plus transparent, participatif et inclusif, aligné avec la vision "Sénégal 2050" portée par ses dirigeants.

Dans un monde où l'information est devenue une ressource stratégique, ce type d'outil illustre comment l'intelligence artificielle peut être mise au service du bien commun et du développement démocratique.

Annexes

A. Guide d'installation

1. Cloner le dépôt :

```
git clone https://github.com/tokosel/Visionsenegal2050.git
cd Visionsenegal2050
```

2. Créer un environnement virtuel :

```
# Avec Python
python -m venv env
# Activation
.\env\Scripts\activate

# Ou avec Conda
conda create --name env python==3.10
conda activate env
```

3. Installer les dépendances :

```
pip install -r requirements.txt
```

4. Configurer les variables d'environnement :

- Créer un fichier `.env`
- Ajouter la clé API Google GEMINI

B. Utilisation

1. Lancer le pipeline d'ingestion des documents :

```
python pipeline.py
```

2. Démarrer l'application Streamlit :

```
streamlit run app.py
```

3. Accéder à l'interface via le navigateur :

```
http://localhost:8501
```

4. Pour utiliser l'API :

```
# Lancement de l'API
uvicorn api:app --host 0.0.0.0 --port 8000 --reload

# Interface de documentation
http://localhost:8000/docs
```