


Projet DeepLearning : Génération de Dialogues avec BiLSTM

Objectif du Projet

L'objectif est de **construire un modèle basé sur BiLSTM** qui peut **générer des dialogues de films** à partir du **Cornell Movie Dialogs Corpus**. Le modèle pourra apprendre à **imiter le style des répliques**, générer de **nouvelles conversations** et répondre à des phrases comme un personnage de film ! 

◇ **Partie 1 : Comprendre le Contexte**

1 Pourquoi Utiliser les BiLSTM ?

Les dialogues sont **des séquences de phrases** où le contexte est important. Un **réseau BiLSTM** est capable de **mémoriser des dépendances** entre les mots et de générer des réponses cohérentes.

2 Dataset Utilisé : Cornell Movie Dialogs

 **Cornell Movie Dialogs Corpus** contient :

- ✓ **220 000 lignes de dialogue de 617 films.**
- ✓ **9 000 personnages** avec leur nom et leur film.
- ✓ **Conversations groupées** (qui parle à qui).

 **Lien du dataset** : [Cornell Movie-Dialogs Corpus](#)

◇ **Partie 2 : Pipeline du Projet**

Étape 1 : Préparation des Données

1. **Télécharger le dataset et extraire les dialogues.**
2. **Nettoyer les textes :**
 - Supprimer la ponctuation inutile.
 - Convertir en minuscules.
 - Supprimer les caractères spéciaux.
3. **Créer des paires question-réponse :**
 - Exemple :
 - **Question** : *"Tu es prêt pour l'aventure ?"*

- **Réponse :** *"Toujours ! Où allons-nous ?"*

Étape 2 : Tokenisation et Encodage des Données

- Convertir les **mots en indices** grâce à `Tokenizer()`.
- Limiter la séquence à **40 mots maximum**.
- Ajouter du **padding** (`pad_sequences`) pour uniformiser la longueur.

Étape 3 : Création du Modèle BiLSTM

Nous allons créer un **modèle séquentiel avec Keras** comprenant :

- ✓ Une **couche Embedding** pour convertir les mots en vecteurs.
- ✓ Deux **couches LSTM bidirectionnelles** pour capturer le contexte.
- ✓ Une **couche Dense Softmax** pour prédire le mot suivant.

Étape 4 : Entraînement du Modèle

Nous allons entraîner notre modèle avec un **batch de 128 et 100 epochs (ou plus si besoin)**.

Étape 5 : Génération de Dialogues

Une fois le modèle entraîné, il pourra **générer des dialogues** réalistes à partir d'une phrase donnée !

Extensions Possibles

- ✓ Ajouter un **chatbot interactif** avec Flask ou Streamlit.
 - ✓ Entraîner un modèle spécifique sur **un seul film** (ex: dialogues de "Matrix").
 - ✓ Tester avec **une température variable** pour plus de diversité.
-