**Documentation du Script d'Entraînement et d'Évaluation de SimpleT5**

Ce script Python utilise la bibliothèque SimpleT5 pour entraîner un modèle T5 (Text-to-Text Transfer Transformer) à partir d'un ensemble de données et évalue ses performances en générant des résumés textuels. Voici les étapes principales du script :

**Installation de SimpleT5:**

- Le script commence par installer la bibliothèque SimpleT5 à l'aide de la commande `!pip install simplet5`.

**Importation des Bibliothèques et des Données:**

- Les bibliothèques nécessaires, telles que pandas et scikit-learn, sont importées.

- Les données d'entraînement et de test sont chargées à partir d'un fichier CSV.

**Préparation des Données:**

-Les colonnes du DataFrame sont renommées en "source\_text" et "target\_text", conformément aux exigences de SimpleT5.

- Un préfixe "summarize:" est ajouté aux textes source pour identifier le type de tâche.

- Les données sont divisées en ensembles d'entraînement et de test.

**Entraînement du Modèle:**

- Un modèle SimpleT5 est instancié et chargé à partir d'un modèle pré-entraîné.

- Le modèle est entraîné sur les données d'entraînement avec des paramètres tels que la longueur maximale des tokens source et cible, le nombre d'époques, et la taille de lot (batch size).

**Génération des titres:**

- Le modèle est utilisé pour générer des titres pour un exemple de texte source de test.

**Évaluation des Performances:**

- Les scores ROUGE (ROUGE-1 et ROUGE-L) sont calculés pour évaluer la qualité des résumés générés.

- Les scores ROUGE sont calculés en comparant les résumés générés aux titres de référence (actual titles) du jeu de test.

**Prérequis:**

Assurez-vous d'avoir installé les bibliothèques nécessaires, y compris SimpleT5, et d'avoir préparé un fichier CSV contenant les données d'entraînement et de test.

**Personnalisation:**

- Vous pouvez personnaliser les paramètres d'entraînement du modèle, tels que la longueur maximale des tokens et le nombre d'époques.

- Vous pouvez également personnaliser les métriques d'évaluation, telles que les scores ROUGE.

Assurez-vous d'avoir suffisamment de puissance de calcul pour l'entraînement du modèle, en particulier si vous utilisez le GPU.

N'oubliez pas de remplacer le chemin du fichier CSV de données (`"/content/drive/MyDrive/Final\_dataset.csv"`) par le vôtre, et assurez-vous d'avoir accès aux fichiers et ressources nécessaires pour exécuter le script.

**AUTEURS :**

* Msalek Aicha
* Hajji chakir Zouhair
* Chebbab Aya
* Houssam Meryem
* Rakhiss Bouchra
* Elouafi Abderrahmane