

Documentation: Modèle GPT-2 Fine-tuné avec des Tweets sur le COVID-19

Introduction

Vit'IA est un chatbot intelligent conçu pour fournir des informations précises et actualisées sur le COVID-19. Il exploite un modèle de langage GPT-2 fine-tuné sur des tweets relatifs à la pandémie de coronavirus pour répondre efficacement aux questions des utilisateurs. Cette documentation présente les fonctionnalités de Vit'IA avec les capacités spécifiques du Modèle GPT-2 afin d'offrir une solution complète pour l'interaction et l'information liées au COVID-19.

Exigences Système

Pour utiliser cet outil, il faut avoir accès à un environnement supportant Python (par exemple, Google Colab) ainsi que l'accès à Google Drive pour le stockage et la récupération des données. De plus, TensorFlow et Keras sont nécessaires pour la création et l'entraînement du modèle.

Configuration et Préparation

Nous avons utilisé Google Collab pour notre environnement d'exécution et nous l'avons connecté à un drive pour avoir notre jeu de données à disposition. Il est nécessaire d'installer les bibliothèques requises pour le traitement de texte et le machine learning. Les tweets sont prétraités à partir de fichiers CSV, convertis en ASCII, et nettoyés pour supprimer les emails, les URLs et les caractères spéciaux. Les données ont ensuite été converties en une liste afin d'être utilisées dans le modèle.

Initialisation et Entraînement

Le modèle GPT-2 est initialisé avec la bibliothèque KerasNLP et configuré avec un préprocesseur afin de préparer nos données à l'entraînement. De plus, nous avons optimisé les hyperparamètres pour que l'entraînement de notre modèle soit le plus précis possible. L'accuracy est la mesure sur laquelle nous nous sommes basés pour connaître le degré de précision de notre modèle.

Utilisation et Interface

Nous utilisons le modèle entraîné pour générer des réponses précises aux questions liées au COVID-19. Gradio est utilisée pour créer une interface web interactive conviviale pour le chatbot. L'interface et le CSS sont personnalisés pour une expérience utilisateur optimale.

Cas d'Utilisation et Limitations

Le chatbot peut répondre à diverses questions telles que les symptômes, les conseils de prévention et les mesures sanitaires. Cependant, il est important de noter que le Modèle GPT-2 Fine-tuné est spécifique au domaine du COVID-19 et que les textes générés peuvent parfois manquer de cohérence ou de pertinence.

Références

- Modèle GPT-2 : [GitHub - SAE_Chatbot](https://github.com/ClaudeZhang7/SAE_Chatbot)
- Hugging Face de notre modèle et de l'interface : [Huggingface - chatbot-gpt2-covid](<https://huggingface.co/Dofla/chatbot-gpt2-covid>)
- Les données d'entraînement : [Kaggle - covid-19-nlp-text-classification](<https://www.kaggle.com/datasets/datatattle/covid-19-nlp-text-classification>)