

Sujet 24

Proposition d'encadrement de projet

A préciser: *Projet de Découverte de la Recherche*

Laboratoire : CReSTIC / LORIA

Équipe : Num&Soc / Synalp

Encadrant : Bart LAMIROY

Bureau :

Téléphone : 06 02 14 69 73

E-mail : **Bart.Lamiroy@univ-reims.fr**

Présentation du sujet (1 page max)

Titre : Écouter parler une IA ou comment mesurer la diversité expressive des modèles de langage automatiques

Description :

Certaines classes de modèles de langage automatiques (dont les plus médiatisés actuellement sont ChatGPT et Gemini) ne sont finalement que de plus ou moins grandes machines d'inférence statistique. En fonction de la taille sous-jacente des réseaux de neurones qui les supportent, et la quantité/qualité des données sur lesquels ils ont été entraînés, leur niveau d'expressivité et la variabilité des réponses qu'ils peuvent produire peut varier grandement.

Dans ce projet, nous proposons (à une échelle modeste) d'étudier des façons d'évaluer la qualité expressive et la variabilité des textes produits par ce type de modèle. Nous utiliserons pour cela une partie de la plateforme Lamuse (<https://lamuse.univ-reims.fr>) qui génère des descriptifs courts à partir d'images d'œuvres d'art.

Pour cela nous explorerons des métriques classiques utilisées en traitement automatique de la langue naturelle comme ROUGE et BLEU et des techniques automatiques de regroupement, d'analyse de données et de classification.

Mots clés : modèles de langage, génération de textes, évaluation

Environnement de travail (matériel, logiciel) :

Le travail se fera en Python en utilisant les bibliothèques classiques de modélisation, entraînement et exécution de réseaux de neurones artificiels, Tensorflow et Keras. En cas de nécessité, des accès à des plateformes de calcul hébergeant des GPU pourra être mis en place. L'encadrant étant bi-localisé entre Nancy et Reims, une partie du suivi se fera à distance.

Références :

[1] « *La muse: Leveraging Artificial Intelligence for Sparking Inspiration* », B. Lamiroy and E. Potier, <https://hal.archives-ouvertes.fr/hal-03470467>