English version below.

Fiche projet en R & D
Sciences des données - IA

Renseignements généraux

OFFRE:

- Étudiant au doctorat ou à la maîtrise
- 4 mois **temps partiel** ou temps plein (Environ 225 h)
- 10 000$ en bourse

Été 2023

Nom et prénom du responsable en entreprise : Nabil Tayeb

Courriel du responsable en entreprise : **nabil@dng.ai** + en CC. entrepreneuriat@ivado.ca

Titre du stage : **Développement et optimisation d'un modèle d'IA multimodal pour le suivi précis d'instructions dans divers domaines d'application**

Mots clés: LLM - ChatGPT - GPT4 - Reinforcement Learning - Machine Learning - Large Language Models - Alpaca - LLama

Description de l'entreprise:

Draft & Goal est une société qui propose un flux de production de contenu automatisé par l'IA. Leur flux de travail permet l'analyse du contenu, l'idéation et la génération d'IA avec un apport externe minimum.

Description du projet:

L'objectif de ce projet de recherche est de développer un modèle de suivi d'instructions utilisant des données textuelles, des images, des sons, des voix et des vidéos. L'objectif principal de ce projet est de construire un modèle capable de suivre des instructions avec précision et de fournir un résultat correct.

Méthodologie :

- Le projet de recherche consistera à entraîner le modèle à l'aide d'un modèle préexistant à source ouverte tel que le modèle Llama. Le modèle sera affiné sur les données de suivi des instructions générées par les techniques récentes sur l'auto-instruction.
- Les performances du modèle seront évaluées à l'aide d'un ensemble de données de test afin de garantir sa précision et son efficacité.

Résultats attendus :
Les résultats attendus de ce projet de recherche sont les suivants
1. Développement d'un modèle de suivi des instructions utilisant diverses formes de données. (texte, image, vidéo)
2. Amélioration de la précision et de l'efficacité du modèle par le biais de fine tuning et d'une évaluation.
3. Optimisation du modèle pour pouvoir le faire fonctionner sur un seul GPU (distillation, hyperparamètres, benchmark …)
4. Réduction du phénomène d'hallucination grâce à des techniques de Retrieval-Augmented-Generation
5. Mise en place et utilisation d'une base de donnée de vecteurs pour maintenir le modèle à jour et citer les sources
6. Création d'un système de traçabilité des demandes et réponses pour évaluer la pertinence des demandes et des réponses avec un niveau de confiance (modèle d'évaluation, mesure de la qualité des réponses)

En conclusion, ce projet de recherche vise à développer un modèle de suivi des instructions qui peut suivre avec précision les instructions en utilisant diverses formes de données. Les résultats du projet contribueront au développement de modèles d'IA plus avancés et aideront à l'automatisation de diverses tâches.

Les compétences recherchés:

(il n'est pas nécessaire de posséder toutes les compétences.)

- Connaissance approfondie des concepts d'intelligence artificielle et de l'apprentissage automatique
- Expérience en programmation avec des langages tels que Python
- Maîtrise des bibliothèques d'apprentissage automatique et de deep learning, telles que TensorFlow, Keras, PyTorch, ou Scikit-learn
- Expérience en traitement de données multimédia (texte, images, sons, voix et vidéos) et en extraction de caractéristiques pertinentes
- Compréhension des réseaux de neurones, en particulier des architectures de réseaux de neurones utilisées pour le traitement des séquences (par exemple, les LSTM) et du traitement multimodal (par exemple, les réseaux de neurones convolutionnels)
- Connaissance des méthodes de prétraitement des données, telles que la normalisation, le redimensionnement et l'augmentation des données
- Familiarité avec les techniques de réduction de la dimensionnalité, telles que l'analyse en composantes principales (ACP) et l'autoencodage
- Capacité à concevoir, mettre en œuvre et évaluer des expériences pour mesurer la performance des modèles d'apprentissage automatique
- Connaissance des méthodes de validation croisée et de sélection de modèles pour éviter le surapprentissage et le sous-apprentissage
- Compétences en optimisation et en réglage des hyperparamètres pour améliorer les performances des modèles
- Capacité à interpréter les résultats et à communiquer les résultats de manière claire et concise
- Connaissance des normes éthiques en matière de recherche en IA et des questions de confidentialité et de sécurité des données

R & D Project Sheet
Data Science - AI

Informations

OFFER:

- Ph.D. or Master's student.
- 4 months part time or full time (Approximately 225 h)
- 10 000$ in scholarship

Summer 2023

Person in charge at the company: Nabil Tayeb

Emails: **nabil@dng.ai** + en CC. entrepreneuriat@ivado.ca

Title of the project: **- Development and optimization of a multimodal AI model for accurate instruction tracking in various application areas**

Key words:
LLM - ChatGPT - GPT4 - Reinforcement Learning - Machine Learning - Large Language Models - Alpaca - LLama

Company description:

Draft & Goal is a company that offers an AI-automated content production workflow. Their workflow enables content analysis, ideation, and AI generation with minimal external input.

Project description :

The objective of this research project is to develop an instruction-tracking model using textual data, images, sounds, voices, and videos. The main objective of this project is to build a model capable of following instructions accurately and providing a correct result.

Methodology:
The research project will involve training the model using a pre-existing open-source model such as the Llama model. The model will be refined on instruction tracking data generated by recent techniques on self-instruction.
The model's performance will be evaluated using a test dataset to ensure its accuracy and effectiveness.

Expected Results:
The expected outcomes of this research project are.
1. Development of an instruction tracking model using various forms of data. (text, image, video)
2. Improvement of the accuracy and efficiency of the model through fine-tuning and evaluation.
3. Optimization of the model to run on a single GPU (distillation, hyperparameters, benchmark ...)
4. Reduction of the hallucination phenomenon thanks to Retrieval-Augmented-Generation techniques
5. Implementation and use of a vector database to keep the model up to date and to quote the sources
6. Creation of a traceability system for requests and responses to evaluate the relevance of requests and responses with a level of confidence (evaluation model, measurement of the quality of responses)

In conclusion, this research project aims to develop an instruction-tracking model that can accurately track instructions using various forms of data. The project results will contribute to the development of more advanced AI models and help in the automation of various tasks.

Skills :

(not all skills are required).

- Strong knowledge of artificial intelligence concepts and machine learning
- Experience programming with languages such as Python
- Proficiency with machine learning and deep learning libraries, such as TensorFlow, Keras, PyTorch, or Scikit-learn
- Experience in processing multimedia data (text, images, sound, voice, and video) and extracting relevant features
- Understanding of neural networks, particularly neural network architectures used for sequence processing (e.g., LSTMs) and multimodal processing (e.g., convolutional neural networks)
- Familiarity with data pre-processing methods, such as normalization, resizing, and data augmentation
- Familiarity with dimensionality reduction techniques, such as principal component analysis (PCA) and autoencoding
- Ability to design, implement, and evaluate experiments to measure the performance of machine learning models
- Knowledge of cross-validation and model selection methods to avoid overlearning and underlearning
- Skills in optimizing and tuning hyperparameters to improve model performance
- Ability to interpret results and communicate results in a clear and concise manner
- Knowledge of ethical standards in AI research and data privacy and security issues