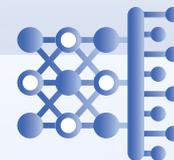


Automatic Multi-modal Deep Learning Analysis System



Principal Investigator

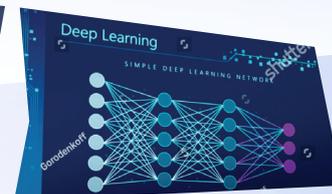
Prof Yu Leung Ho Philip

An automated system that assesses the accuracy and relevance of users' input (e.g. textual descriptions) paired with various modes of communication (e.g. images, videos, audio, etc.), and provides immediate, high-quality personalised feedback to users at different levels to progressively improve their expertise by leveraging a finetuned Multimodal Large Language Model (MLLM), deep learning and contrastive learning techniques and generative AI technology.



Features and Advantages

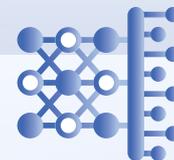
- 1 Evaluates domain-specific datasets through customisation, such as medical imaging, education, and large datasets
- 2 Can be adapted for industry-scale commercial use based on proven deep learning techniques
- 3 Automatic scoring based on description accuracy and relevance
- 4 Uncovers quality issues that were not previously obvious in large datasets, improving dataset management
- 5 Strong agreement between system feedback and human evaluation, effectively assisting or even replacing manual quality checks



Future Development

- ▶ Adapt for educational and training purposes, such as automated grading of image description tasks in language learning or testing environments, e.g. assessing learners' ability to describe images accurately, training of pathology or medical professionals, and in other fields
- ▶ Automatically assess and maintain the quality of large datasets used for image captioning, visual question answering, and image-text retrieval
- ▶ Verify the quality of the generated descriptions on platforms that rely on large-scale human annotation without needing manual checks
- ▶ Improve AI model performance in real-world tasks

Système d'analyse automatique multimodal d'apprentissage profond



Chercheur principal
 Prof Yu Leung Ho Philip

Un système automatisé qui évalue l'exactitude et la pertinence des entrées des utilisateurs (par exemple, des descriptions textuelles) associées à divers modes de communication (par exemple, des images, des vidéos, de l'audio, etc.), et fournit un retour d'information personnalisé immédiat et de haute qualité aux utilisateurs à différents niveaux afin d'améliorer progressivement leur expertise en tirant parti d'un modèle MLLM (Multimodal Large Language Model) affiné, de techniques d'apprentissage profond et d'apprentissage contrastif et de la technologie d'IA générative.



Caractéristiques et avantages

- 1 Évalue les ensembles de données spécifiques à un domaine grâce à la personnalisation, tels que l'imagerie médicale, l'éducation et les grands ensembles de données
- 2 Peut être adapté pour une utilisation commerciale à l'échelle de l'industrie sur la base de techniques d'apprentissage profond éprouvées
- 3 Notation automatique basée sur l'exactitude et la pertinence de la description
- 4 Révèle des problèmes de qualité qui n'étaient pas évidents auparavant dans les grands ensembles de données, améliorant ainsi la gestion des ensembles de données
- 5 Une forte concordance entre le retour d'information du système et l'évaluation humaine, ce qui permet d'assister, voire de remplacer, les contrôles de qualité manuels

Développement futur

- ▶ S'adapter à des fins d'enseignement et de formation, telles que la notation automatisée des tâches de description d'images dans des environnements d'apprentissage des langues ou de tests, par exemple l'évaluation de la capacité des apprenants à décrire des images avec précision, la formation de professionnels de la pathologie ou de la santé, et dans d'autres domaines
- ▶ Évaluer et maintenir automatiquement la qualité des grands ensembles de données utilisés pour le sous-titrage des images, les réponses aux questions visuelles et la récupération d'images-textes
- ▶ Vérifier la qualité des descriptions générées sur des plateformes qui s'appuient sur l'annotation humaine à grande échelle sans avoir besoin de vérifications manuelles
- ▶ Améliorer les performances des modèles d'IA dans des tâches réelles