

Du **30 SEP.**

11h00

LISN Site Belvédère

SÉMINAIRES

Developing the Framed Multi30k (FM30k) multimodal-multilingual dataset 2/2

Akram Elbouanani, Université Paris-Saclay, CEA, List

Orateur : Akram Elbouanani

AJOUTER À MON CALENDRIER

Résumé

Les modèles de langage transforment la manière dont on analyse les contenus politiques, mais ils peuvent aussi refléter des biais implicites qui influencent leurs réponses. Nous verrons comment, pour une tâche simple, le seul fait de changer le nom d'un politicien dans une phrase peut faire varier le jugement du modèle.

Pour aller plus loin, nous explorerons ces biais à travers plusieurs langues (anglais, français, arabe, russe...) et différents modèles (Mistral, Meta, OpenAI...). Nous essayerons de comprendre comment des facteurs comme le pays d'origine, l'orientation politique, le genre ou la langue d'entraînement influencent ces décalages, avec des biais parfois surprenants.

Nous analyserons les différences entre langues, questionnerons l'origine de ces biais, et réfléchirons à leurs conséquences pour les futurs développements.

Enfin, nous proposerons une méthode simple pouvant déjà aider à limiter ces effets, tout en discutant des défis éthiques majeurs liés à l'utilisation responsable des LLMs dans des contextes politiques de plus en plus polarisés

Référence

<https://arxiv.org/pdf/2505.19776>



LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE

Site Belvédère : Campus Universitaire bât.507 - Rue du Belvédère
91405 - Orsay

Site Plaine : 1 Rue Raimond Castaing et Rue René Thom (Bât Digitéo)
91190 - Gif-sur-Yvette