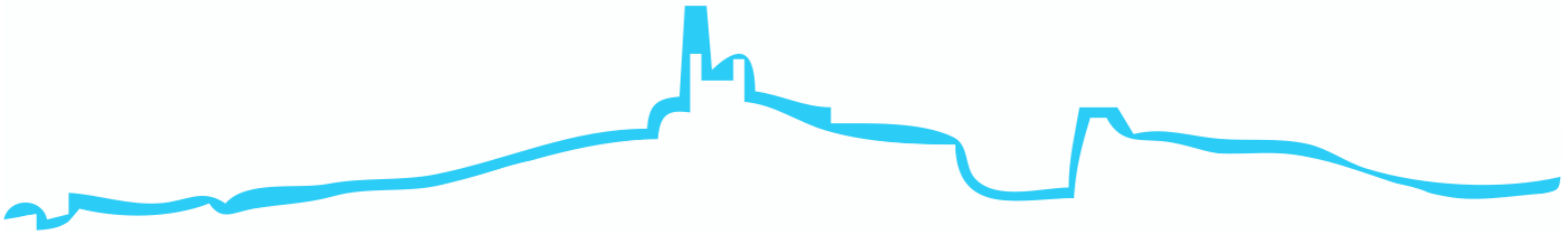

CORIA-TALN 2025

20e Conférence en Recherche d'Information et Applications (CORIA)
32ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN)
27ème Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL)
Les 18e Rencontres Jeunes Chercheurs en RI (RJCRI)
(CORIA-TALN) [1]

Actes de CORIA-TALN-RJCRI-RECITAL 2025.

Actes de l'atelier Ethic and Alignment of (Large) Language Models 2025 (EALM)

Frédéric BECHET, Adrian-Gabriel CHIFU, Karen PINEL-SAUVAGNAT, Benoit FAVRE, Eliot MAES, Diana NURBAKOVA (Éds.)

Marseille, France, 30 juin au 4 juillet 2025

1. https://coria-taln-2025.lis-lab.fr

Avec le soutien de

Préface

L'atelier EALM est organisé par les membres du projet ANR Diké, une collaboration entre Naver Labs Europe, le laboratoire ERIC de l'université Lyon 2 et le laboratoire Hubert Curien de l'université Jean Monnet à Saint-Étienne. Ce projet explore l'intersection entre équité et compression des grands modèles de langage, afin de relever deux défis majeurs : garantir l'équité des modèles et proposer des solutions pour contrer l'augmentation de leur taille, qui peut constituer un frein à leur déploiement.

L'atelier EALM 2025 vise à étudier l'éthique et l'alignement au sens large des (grands) modèles de langage. Cet événement se tiendra sur une demi-journée, en anglais, et comprendra deux présentations invitées ainsi qu'un appel à contribution, permettant de présenter des travaux préliminaires sur ces thématiques. La journée se déroulera le 30 Juin 2025 sur une demi-journée.

Cet atelier proposera des discussions approfondies sur plusieurs thématiques, notamment :
- Alignement des modèles de langage
- Étude de l'équité des LLMs, jeux de données, méthode de réduction de biais
- Considérations éthiques dans le développement des modèles de langage
- Frugalité de l'entraînement et du déploiement des grands modèles de langage
- Impact sociétal des modèles de langage
- Considérations interculturelles des LLMs
- Élicitation des préférences humaines
- LLMs inclusifs

En tant que conférencière invitée, nous aurons Daryna Dementieva, chercheuse postdoctorale au sein du groupe de recherche en informatique sociale (Social Computing Research Group) à l'Université technique de Munich. Nous présenterons ensuite les travaux des trois dernières années réalisés dans le cadre de l'ANR Diké. Trois articles originaux acceptés seront également présentés.

En termes des soumission, 4 articles pour la conférence principale ont été soumis, dont respectivement 3 ont été acceptés pour une présentation orale

Nous remercions chaleureusement toutes les personnes ayant contribué à la réussite de l'organisation de cet atelier : les auteurs, les relecteurs, les comités scientifiques. Un merci particulier aux équipes de la conférence CORIA-TALN 2025, notamment aux membres du comité d'organisation qui nous ont apporté leur soutien pour la création de cet atelier, Carlos Ramisch, Ismail Badache et Benoit Favre

Antoine Gourru, Julien Velcin, Caroline Brun, Vassilina Nikoulina, Thibaud Leteno et Irina Proskurina

Comités

Comité de Programme

— Antoine Gourru, Université Jean Monnet, Laboratoire Hubert Curien
— Julien Velcin, Université de Lyon, Lyon 2, Laboratoire ERIC
— Caroline Brun, Naver Lab, Grenoble
— Vassilina Nikoulina, Naver Lab, Grenoble
— Thibaud Leteno, Université Jean Monnet, Laboratoire Hubert Curien
— Irina Proskurina, Université de Lyon, Lyon 2, Laboratoire ERIC

Comité de Relecture

— Antoine Gourru, Université Jean Monnet, Laboratoire Hubert Curien
— Julien Velcin, Université de Lyon, Lyon 2, Laboratoire ERIC
— Caroline Brun, Naver Lab, Grenoble
— Vassilina Nikoulina, Naver Lab, Grenoble
— Thibaud Leteno, Université Jean Monnet, Laboratoire Hubert Curien
— Irina Proskurina, Université de Lyon, Lyon 2, Laboratoire ERIC

Table des matières