

InExtenso : Journée rouennaise

Karën Fort

karen.fort@loria.fr

CHU de Rouen, 3 déc. 2024

Bienvenue !

▶ LORIA

- ▶ Karën Fort, UL/LORIA
- ▶ Clémentine Bleuze (doctorante), LORIA
- ▶ Amandine Lecomte (doctorante/ingénieure), LORIA

▶ LISN

- ▶ Aurélie Névéol
- ▶ Pierre Zweigenbaum
- ▶ Ayoub Hammal (doctorant) : à distance
- ▶ Caio Corro (Rennes) : à distance, une partie de la journée
- ▶ Nona Naderi : à distance (?)

▶ CHU de Rouen

- ▶ Stefan Darmoni
- ▶ Romain Lelong
- ▶ Francesco Monti
- ▶ Hélène Cieslik, cheffe de projet

Rappel de l'organisation de la journée

9h15 - 12h15

- ▶ Point sur le projet (KF)
- ▶ Aurélie : papier sur les biais dans la génération (soumis)
- ▶ Clémentine Bleuze : activités de recherche sur l'état de l'art et journée d'hier
- ▶ (visio) Ayoub Hammal : approches pour modifier les productions des LLMs "brutes" et pistes de recherche

Déjeuner : 12h15 - 13h45

13h45-16h

- ▶ Amandine Lecompte, Pôle psy : point sur les données, les besoins, etc
- ▶ Rouen (suggestion) : point sur les données et sur la partie psy
- ▶ Discussions sur les collaborations

Du mouvement...

Arrivées :

- ▶ Clémentine Bleuze : doctorante, dir. A. Névéol et K. Fort
- ▶ Ayoub Hammal : doctorant, dir. PZ, C. Corro et M. Couceiro
- ▶ autre ?

Changements :

- ▶ C. Corro : mutation MCF à Rennes
- ▶ K. Fort : PR à Nancy
- ▶ M. Couceiro : PR à Lisbonne

Premiers livrables réalisés

- ▶ DMP (plan de gestion des données) provisoire : question des données de Rouen ?
- ▶ site Web : <https://anr-inextenso.loria.fr/>

Collaborations : réunion formation avec les autres TSIA 2023

19 nov. 2024, à l'ISIR

- ▶ 14 présents (dont 7 doctorant·e·s)
- ▶ Pantagruel, MALADES, Guidance (LLM4All excusés)
- ▶ Matinée sur les impacts env., aprem sur les stéréotypes
- ▶ Bcp d'intérêt chez les doctorant·e·s
- ▶ Collaborations plus étroites à venir (à définir, **relancer**)

Publications : évaluation

- ▶ Fanny Ducel, Aurélie Névéol, Karën Fort. Desiderata for Actionable Bias Research. New Perspectives on Bias and Discrimination in Language Technology, Nov 2024, Amsterdam (Pays-Bas), France. hal-04755691
- ▶ Fanny Ducel, Aurélie Névéol, Karën Fort. "You'll be a nurse, my son!" Automatically Assessing Gender Biases in Autoregressive Language Models in French and Italian. Language Resources and Evaluation, 2024. hal-04803403
- ▶ Ajinkya Kulkarni, Atharva Kulkarni, Isabel Trancoso, Miguel Couceiro. Unveiling Biases while Embracing Sustainability : Assessing the Dual Challenges of Automatic Speech Recognition Systems. Interspeech 2024, Sep 2024, Kos / Greece, Greece. hal-04610235
- ▶ Fanny Ducel, Aurélie Névéol, Karën Fort. Évaluation automatique des biais de genre dans des modèles de langue auto-régressifs. TALN 2024, Jul 2024, Toulouse, France. hal-04621134
- ▶ Clément Morand, Aurélie Névéol, Anne-Laure Ligozat. MLCA : a tool for Machine Learning Life Cycle Assessment. 2024 International Conference on ICT for Sustainability (ICT4S), Jun 2024, Stockholm, Sweden. hal-04643414

Publications : débiaisage

- ▶ Guilherme Dean Pelegrina, Miguel Couceiro, Leonardo Tomazeli Duarte. A preprocessing Shapley value-based approach to detect relevant and disparity prone features in machine learning. ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT 2024), Fabro Steibel, Meg Young, Ricardo Baeza-Yates, Jun 2024, Rio de Janeiro (BR), Brazil. pp.279-289, 10.1145/3630106.3658905. hal-04568343
- ▶ Ajinkya Kulkarni, Anna Tokareva, Mohammed Rameez Qureshi, Miguel Couceiro. The Balancing Act : Unmasking and Alleviating ASR Biases in Portuguese. EACL 2024 LT-EDI WorkShop, Mar 2024, St. Julians, Malta. hal-04436147

Merci de mettre vos publications InExtenso sur HAL avec le numéro du projet :
ANR-23-IAS1-0004