



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 Accès à l'information basé sur le dialogue et
grands modèles de langage 2025 (DIAG-LLM)

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

Organisateurs



Soutiens académiques



Sponsors privés



Préface

Organisé conjointement par Benjamin Piwowarski (CNRS, ISIR, Sorbonne Université) et Josiane Mothe (Université de Toulouse Jean-Jaurès, IRIT, CNRS) avec le soutien du projet ANR GUIDANCE, l'atelier "Accès à l'information basé sur le dialogue et grands modèles de langage" (DIAG-LLM) vise à rassembler la communauté autour des questions liées à l'accès à l'information basé sur le dialogue et l'intégration de la génération automatique augmentée par la recherche d'information pour améliorer les systèmes de recherche conversationnelle.

Les Grands Modèles de Langage (LLMs) sont en train de transformer l'accès à l'information en facilitant l'interaction en langage naturel. Cependant, les systèmes actuels présentent plusieurs limitations :

- Les LLMs ne sont pas conçus initialement pour la RI, et nécessitent des adaptations pour les interactions lors d'une recherche documentaire.
- Problèmes de véracité et d'hallucination, compromettant la fiabilité des réponses générées.
- Faible généralisation aux langues et domaines à faibles ressources, limitant l'accessibilité et nécessitant une meilleure compréhension des fonctionnements internes.
- Défis liés à l'explicabilité, rendant difficile la justification transparente des résultats.

l'atelier DIAG-LLM porte sur les sujets d'intérêt suivants (sans y être limités) :

- Modèles les plus adaptés à la RI Conversationnelle
 - Combinaison des approches de RI denses et parcimonieuses
 - Apprentissage en continu
- Architectures pour un accès interactif à l'information
- Stratégies de clarification et reformulation en recherche conversationnelle
- Conception de prompts spécialisés pour la génération de requêtes
- Évaluation et collections pour la recherche d'information conversationnelle
 - Réflexion sur les collections et mesures d'évaluation, début de collections
 - Collection de données et outils d'annotation pour l'évaluation
- Explicabilité dans les modèles neuronaux de recherche d'information
- Adaptation aux langues et domaines à faibles ressources
 - Apprentissage zéro-shot et few-shot pour l'adaptation aux nouveaux domaines
 - Gestion de la diversité linguistique et adaptation aux langues peu dotées
- Factualité, biais et véracité
 - Détection et atténuation des hallucinations dans les réponses générées par les LLMs
 - Évaluation de l'équité et des biais dans les systèmes conversationnels
 - Mécanismes de vérification des faits pour l'accès à l'information

Cet atelier s'adresse aux chercheurs et professionnels des domaines suivants :

- Recherche d'Information (RI) et Traitement du Langage Naturel (TAL)
- Intelligence Artificielle Conversationnelle et Systèmes de Dialogue
- Explicabilité et Fiabilité de l'IA
- Traitement des Langues Faiblement Dotées

L'atelier débutera par des communications sélectionnées suite à un appel ouvert :

- Introduction à l'atelier - Y. Chasseray, B. Piwowarski
- Contribution à DIAG-LLM : une revue de littérature - R. Bourgeois, S. Fournier et A. Chifu
- Explicabilité par Perturbations pour les Systèmes RAG - Y. Zhou, P. Mulhem et D. Schwab
- Few-Shot prompting strategies : Random or Selective - R. Kumar
- TURNER - ModernSPLADE : is bert-base the ONLY base ? - J. Drexal

Elles seront suivies par une table ronde autour du projet GUIDANCE. Enfin, le groupe GUIDANCE présentera les propositions pour la participation au challenge *i – Kat* de TREC (Interactive Knowledge Assistance). L'atelier sera suivi d'un hackathon pour la participation à *i – Kat*.

Nous tenons à exprimer notre profonde reconnaissance envers toutes les personnes qui ont participé à faire vivre l'atelier : les autrices et auteurs des soumissions, les membres du comité de relecture, et le comité d'organisation de CORIA-TALN 2025. Nous remercions également très chaleureusement l'ANR. Son soutien financier via le projet GUIDANCE nous a permis de réunir des intervenants et des participants spécialistes ainsi que de fédérer cette communauté autour d'un sujet d'actualité.

Benjamin Piwowarski
Josiane Mothe
Organisateurs de DIAG-LLM

Comités

Comité de Programme

- Benjamin Piwowarski - CNRS, ISIR, Sorbonne Université
- Josiane Mothe - Université de Toulouse Jean-Jaurès, INSPE, IRIT, CNRS

Comité de Relecture

- Yann Chasseray, IRIT, Toulouse
- Adrian Chifu, LIS, Marseille
- Sébastien Fournier, LIS, Marseille
- Lorraine Goeriot, LIG, Grenoble
- Josiane Mothe, IRIT, Toulouse
- Benjamin Piwowarski, CNRS, Paris
- Yongxin Zhou, LIG, Grenoble

Table des matières

Explicabilité par Perturbations pour les Systèmes RAG	1
<i>Yongxin Zhou, Philippe Mulhem, Didier Schwab</i>	
Représentations conditionnelles entité-centrées pour le raisonnement multi-saut dans les systèmes de question-réponse multi-document	7
<i>Romain Bourgeois, Adrian Chifu, Sébastien Fournier</i>	