UNIVERSITÉ D'ARTOIS

Avis de Soutenance

Monsieur Naïm ES SEBBANI

Informatique et applications

Soutiendra publiquement ses travaux de thèse intitulés
De la représentation sémantique au raisonnement logique dans les modèles de langage

Travaux dirigés par Monsieur Zied BOURAOUI et Monsieur Jan GOES

Soutenance prévue le ***lundi 16 novembre 2026*** à 10h00
Lieu : Université d'Artois, Faculté des Sciences Jean Perrin, rue Jean Souvraz SP 18, 62307 Lens Cedex
Salle : des thèses

Composition du jury proposé

M. Zied BOURAOUI	Université d'Artois	Directeur de thèse
Mme Fatiha SAIS	Paris Saclay University	Examinatrice
M. Pierre GENEVÈS	CNRS - Inria Grenoble Rhône-Alpes	Rapporteur
Mme Lina SOUALMIA	Université de Rouen Normandie	Rapporteure
M. Jan GOES	Université d'Artois	Co-directeur de thèse
M. Jean-François CONDOTTA	Université d'Artois	Examinateur
M. Yakoub SALHI	Université d'Artois	Invité

Résumé :

Découvrir les relations entre entités, fragments de texte et énoncés logiques est un défi central du traitement automatique des langues. Cette thèse étudie la connaissance relationnelle à trois niveaux de granularité, en demandant si les modèles de langue actuels représentent la structure relationnelle et raisonnent véritablement sur celle-ci. Nous traitons d'abord les relations entre entités : nous introduisons SemRelLM, un jeu de plus de 80~000 triplets générés par un grand modèle de langue couvrant un large éventail de relations sémantiques, ainsi que trois encodeurs relationnels contrastifs qui capturent des facettes complémentaires de la sémantique relationnelle et surpassent l'état de l'art en analogie et en classification de relations lexicales. Nous étendons ensuite cette perspective aux documents longs avec le Chunk Prediction Encoder (CPE), un cadre auto-supervisé qui modélise la compréhension d'un document comme une compatibilité entre fragment et contexte, et apporte des gains constants sur des benchmarks juridiques et biomédicaux tout en restant efficace, avec des encodeurs hiérarchiques et à sparse attention. Enfin, nous examinons si les modèles de raisonnement raisonnent réellement sur des relations logiques, via un benchmark diagnostique fondé sur des familles paramétrées de formules 2-CNF, à structure de graphe d'implication contrôlable et perturbations préservant la satisfiabilité. Nos expériences révèlent une fragilité propre à chaque générateur, compatible avec l'hypothèse d'une réutilisation de motifs familiers plutôt que d'un raisonnement structurel invariant à la présentation.