Avis de Soutenance

Monsieur Thomas BAILLEUX

Informatique et applications

Soutiendra publiquement ses travaux de thèse intitulés
Apprendre et raisonner à partir d'explications en langues naturelles

Travaux dirigés par Monsieur Zied BOURAOUI

Soutenance prévue le ***mercredi 25 novembre 2026*** à 10h00
Lieu : 13 Rue Jean Souvraz, Bâtiment A, 62300, Lens
Salle : des Thèses

Composition du jury proposé

M. Zied BOURAOUI	Université d'Artois	Directeur de thèse
Mme Marie-Jeanne LESOT	Sorbonne Université	Rapporteure
Mme Céline HUDELOT	CentraleSupélec	Rapporteure
M. Gilles AUDEMARD	Université d'Artois	Examinateur
M. Nasredine SEMMAR	CEA - Paris-Saclay	Examinateur
M. Emmanuel LONCA	Université d'Artois	Invité
M. Tanmoy MUKHERJEE	Université d'Artois	Invité

Résumé :

La langue naturelle sert à la fois d'interface avec les systèmes d'intelligence artificielle et de support d'explications. La fluidité d'une explication ne garantit cependant ni sa qualité ni sa fidélité au comportement du modèle. Cette thèse étudie comment rendre les explications et les représentations intermédiaires explicites afin de les comparer, de les analyser et de les manipuler. Trois contributions sont proposées. La première formule l'évaluation de la qualité des explications comme un problème de classement entre candidats et étudie des objectifs ponctuels, par paires et par listes, ainsi que leur utilisation comme signaux de récompense. La deuxième, CLARITY, relie des extraits justificatifs à des représentations conceptuelles utilisées pour la classification. Dans sa configuration sans voie de contournement, la prédiction suit cette chaîne, rendant explicite le passage des éléments linguistiques sélectionnés aux variables intermédiaires du modèle. La troisième, CREDENCE, enrichit les modèles à goulot d'étranglement conceptuel avec la dispersion entre prédicteurs conceptuels et une sortie supervisée décrivant le caractère indéterminé de certaines annotations conceptuelles. Ces travaux fournissent des formulations et des architectures pour étudier la qualité relative des explications, le rôle des représentations conceptuelles dans la prédiction et les associations entre signaux d'incertitude et erreurs. Leur évaluation distingue propriétés architecturales, performance prédictive et validation sémantique ; certaines comparaisons quantitatives et les bénéfices des interventions restent à consolider. Cette thèse contribue ainsi à une approche où les explications et les concepts deviennent des composantes explicites du calcul, accessibles à l'analyse et à une évaluation de leurs propriétés.