



AVIS DE PRESENTATION DE THESE EN SOUTENANCE POUR L'OBTENTION DU DIPLOME NATIONAL DE DOCTEUR

Monsieur Souhail BAKKALI

Présentera ses travaux intitulés :

« Apprentissage Inter-modal et Unifié de la Vision et du langage pour la Compréhension Multimodale des Documents »

Spécialité : Informatique et Applications

Le 14 décembre 2022 à 12h00

Lieu :

**La Rochelle Université
Pôle Communication, Multimédia et Réseaux
Amphithéâtre Michel Crépeau
44 Av. Albert Einstein
17000 LA ROCHELLE**

Composition du jury :

**Mme BARNEY SMITH H. Elisa
M. COUSTATY Mickael
M. FISHER Andreas**

M. JAWAHAR C.V

**Mme KAVALLIERATOU Ergina
M. MARINAI Simone
M. RUSINOL Marçal (*invité*)
M. TERRADES Oriol Ramos
M. ZUHENG Ming (*invité*)**

Professeure, Boise State University Idaho (USA)

Maître de conférences, La Rochelle Université

**Professeur, University of Applied Sciences and Arts Western
Switzerland (Suisse)**

**Professeur, International Institute of Information Technology
Hyderabad (Inde)**

Professeure, University of the Aegean, (Grèce)

Professeur, Università degli Studi di Firenze (Italie)

Maître de conférences, Université Autonome de Barcelone (Espagne)

Maître de conférences, HDR, Université Autonome de Barcelone (Espagne)

Maître de conférences, Université Sorbonne Paris Nord

Résumé :

Les modèles développés dans cette thèse sont le résultat d'un processus itératif d'analyse et de synthèse entre les théories existantes et nos études réalisées. Plus spécifiquement, nous souhaitons étudier l'apprentissage inter-modal pour la compréhension contextualisée sur les composants des documents à travers le langage et la vision.

Cette thèse porte sur l'avancement de la recherche sur l'apprentissage inter-modal et apporte des contributions sur quatre fronts : (i) proposer une approche inter-modale avec des réseaux profonds pour exploiter conjointement les informations visuelles et textuelles dans un espace de représentation sémantique commun afin d'effectuer et de créer automatiquement des prédictions sur les documents multimodaux; (ii) à étudier des stratégies concurrentielles pour s'attaquer aux tâches de classification de documents intermodaux, de récupération basée sur le contenu et de classification few-shot de documents ; (iii) pour résoudre les problèmes liés aux données comme l'apprentissage lorsque les données ne sont pas annotées, en proposant un réseau qui apprend des représentations génériques à partir d'une collection de documents non étiquetés ; et (iv) à exploiter les paramètres d'apprentissage few-shot lorsque les données ne contiennent que peu d'exemples.