

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

Monsieur Vincent POULAIN D'ANDECY

Présentera ses travaux intitulés :

**« Système à connaissance incrémentale pour la compréhension de document et la
détection de fraude »**

Spécialité : Informatique et applications

Le 5 octobre 2021 à 14h00

Lieu :

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

Composition du jury :

Mme AUSSENAC Nathalie	Directrice de recherche, Institut de recherche en informatique de Toulouse
M. COUASNON Bertrand	Maître de conférences, HDR, Institut des sciences appliquées de Rennes
Mme EGLIN Véronique	Professeure, INSA de Lyon
M. INGOLD Rolf	Professeur, Université de Fribourg
M. LLADÓS Josep	Professeur, Université Autonome de Barcelone
M. OGIER Jean-Marc	Professeur, La Rochelle Université

Résumé :

Le Document Understanding est la discipline de l'Intelligence Artificielle qui dote les machines du pouvoir de Lecture. Cette capacité sous-entend de comprendre dans une vision globale l'objet du document, sa classe, et dans une vision locale, des informations précises, des entités. Un double défi est de réussir ces opérations dans plus de 90% des cas tout en éduquant la machine avec le moins d'effort humain possible. Cette thèse défend la possibilité de relever ces défis avec des méthodes à apprentissage incrémental. Nos propositions permettent d'éduquer efficacement et itérativement une machine avec quelques exemples de document. Pour la classification, nous démontrons (1) la possibilité de découvrir itérativement des descripteurs textuels, (2) l'intérêt de tenir compte de l'ordre du discours et (3) l'intérêt d'intégrer dans le modèle de donnée incrémental une mémoire épisodique de quelques *Souvenirs* d'échantillon. Pour l'extraction d'entité, nous démontrons un modèle structurel itératif à partir d'un graphe en étoile dont la robustesse est améliorée avec quelques connaissances a priori d'ordre général.

Conscient de l'importance économique et sociétale de la fraude dans les flux documentaires, cette thèse fait également le point sur cette problématique. Notre contribution est modeste en étudiant les catégories de fraude pour ouvrir des perspectives de recherche.

Cette thèse a été conduite dans un cadre atypique en conjonction avec une activité industrielle à Yooz et des projets collaboratifs, en particulier, les projets FEDER SECURDOC soutenu par la région Nouvelle Aquitaine et Labcom IDEAS soutenu par l'ANR.