

## PROPOSITION DE SUJET DE THESE

**Intitulé : Sélection des données utiles par et pour intelligence artificielle**

Référence : **TIS-DTIS-2025-03**  
(à rappeler dans toute correspondance)

**Début de la thèse : 2025**

**Date limite de candidature : 2025**

### Mots clés

**IA, machine learning, frugalité, apprentissage incrémental**

### Profil et compétences recherchées

M2 recherche en machine learning (ex MVA)

### Présentation du projet doctoral, contexte et objectif

Il est aujourd'hui acquis qu'il est possible d'apprendre une fonction d'un espace de données vers un espace de représentation efficace à partir d'un très gros volume de données : c'est par exemple le cas pour les textes avec les grands modèles de langage comme GPT4, ou pour des images naturelles avec le modèle SAM2. Cependant, cette approche n'est pas satisfaisante pour un grand nombre d'application où il n'est pas envisageable de disposer de gros volume de données (comme la défense ou le médical). De plus cette approche n'est pas satisfaisante vis-à-vis des processus cognitifs mis en œuvre dans l'apprentissage humain : nous ne considérons pas toutes les données comme équivalentes durant notre développement. On peut notamment citer les travaux de Karl Friston qui laissent penser que les situations surprenantes (celles que l'on aura eu du mal à prédire) ou stressantes vont recueillir beaucoup plus d'attention que les situations banales.

Dans le cadre, du développement de modèle de représentation, cette réalité biologique doit nous inviter à considérer des approches plus frugales en données, par exemple en s'intéressant à la question « Quelles données seraient réellement nécessaires pour apprendre une représentation ? »

Cette question n'est pas simplement académique : elle a des implications logistiques très importantes dans la conception de corpus d'apprentissage à partir de flottes de capteurs décentralisés : « à partir d'une flotte de véhicules instrumentés, pourrait-on n'enregistrer que les données importantes ? »

Cette question qui croise les problématiques de frugalité, d'apprentissage incrémentale, de détection d'anomalie sera le cœur de la thèse.

### Collaborations envisagées

La problématique de la sauvegarde frugale des données est aujourd'hui prioritaire dans les forces. Cette thèse conduira donc à des collaborations avec des acteurs de la défense AMIAD, STAT, ...

#### Laboratoire d'accueil à l'ONERA

Département : traitement de l'information et système

Lieu (centre ONERA) : Palaiseau

**Contact** : Adrien Chan-Hon-Tong

Email : [adrien.chan\\_hon\\_tong@onera.fr](mailto:adrien.chan_hon_tong@onera.fr)

#### Directeur de thèse

Nom : Chan-Hon-Tong Adrien

Laboratoire : ONERA

Tél. : 01 80 38 65 91

Email : [adrien.chan\\_hon\\_tong@onera.fr](mailto:adrien.chan_hon_tong@onera.fr)

Pour plus d'informations : <https://www.onera.fr/rejoindre-onera/la-formation-par-la-recherche>