



IMT Mines Alès
École Mines-Télécom

LA SCIENCE & LA CRÉATIVITÉ POUR INVENTER UN MONDE DURABLE



Post-Doctorant / Post-Doctorante en Apprentissage Automatique

Etablissement	IMT Mines Alès (Ecole Nationale Supérieure des Mines d'Alès)
Affectation principale	Centre d'enseignement & de recherche en informatique & systèmes (CERIS)
Résidence administrative	Alès (Département du Gard – Région Occitanie)
Type de contrat	CDD 18 mois – Contrat de droit public – Temps plein
Date de prise de poste	01/07/2024

Cette étude fait partie du **projet européen ENFIELD**. ENFIELD vise à établir un **centre d'excellence européen distinctif axé sur l'avancement de la recherche fondamentale en matière d'IA adaptative, soutenable, centrée sur l'humain et digne de confiance**. Ces piliers représentent de nouveaux éléments stratégiques essentiels au développement, au déploiement et à l'acceptation de l'IA en Europe. L'Institut Mines-Télécom (IMT), partenaire du projet à travers sa communauté scientifique Data Analytics and AI, est en charge du workpackage qui regroupe les recherches du projet sur ces quatre piliers scientifiques. Cette étude s'inscrit dans le pilier de l'IA centrée sur l'humain. Plus précisément, le projet de ce post-doctorat s'inscrit dans le sous-thème "HC-AI.3 Interpretable Data Driven Decision Support Systems" du pilier.

Présentation de notre établissement et du Centre CERIS

L'Institut Mines-Télécom

L'institut Mines-Télécom (IMT), grand établissement au sens du code de l'éducation, est un établissement public scientifique, culturel et professionnel (EPSCP) placé sous la tutelle principale des ministres chargés de l'industrie et du numérique. Premier groupe d'écoles d'ingénieurs en France, il fédère 11 écoles d'ingénieur publiques réparties sur le territoire national, qui forment 13 500 ingénieurs et docteurs. L'IMT emploie 4500 personnes et dispose d'un budget annuel de 400M€ dont 40% de ressources propres. L'IMT comporte 2 instituts Carnot, 35 chaires industrielles, produit annuellement 2100 publications de rang A, 60 brevets et réalise 110M€ de recherche contractuelle.



IMT Mines Alès

Créée en 1843, IMT Mines Alès compte à ce jour 1400 élèves (dont 250 étrangers) et 380 personnels. L'école dispose de 3 centres de recherche et d'enseignement de haut niveau scientifique et technologique, qui œuvrent dans les domaines des matériaux et du génie civil (C2MA), de l'environnement et des risques (CREER), de l'intelligence artificielle et du génie industriel et numérique (CERIS). Elle dispose de 12 plateformes technologiques et compte 1600 entreprises partenaires. La personne recrutée sera affectée au Centre d'Enseignement et de Recherche en Informatique et Systèmes (CERIS).

Centre CERIS

Le CERIS se compose de deux équipes de recherche, ISOAR pour *Ingénierie des Systèmes et des Organisations pour les Activités à Risque* et I3A pour *Informatique, Image et Intelligence Artificielle*. Le CERIS gère de même deux départements d'enseignement : 2IA pour *Informatique et Intelligence artificielle* et PRISM pour *PeRformance Industrielle et Systèmes Mécatroniques* ainsi que 2 plateformes technologiques AIHM pour *Alès Imaging and Human Metrology* et la PFM pour *Plateforme Mécatronique*.

Contexte de l'étude

Nous disposons d'un ensemble de techniques d'apprentissage machine qui permettent le développement de systèmes prédictifs performants pour la résolution de nombreuses tâches. Certains de ces systèmes sont intégrés dans des processus décisionnels impliquant des opérateurs humains, et dans lesquels le prédicteur sera a priori considéré comme une aide à la décision, e.g. prise de décisions délicates dans des contextes complexes seulement partiellement appréciés par le système prédictif (visibilité partielle du contexte décisionnel), ou encore du fait de contraintes légales quant à la décision finale qui sera prise. L'opérateur humain tire dans ce cas parti de prédictions fournies par le système pour prendre sa décision.

Afin d'améliorer ce type de collaboration Homme-Machine, il peut être souhaitable de fournir à l'opérateur humain des informations sur le comportement global du prédicteur, voire d'enrichir les prédictions fournies par des informations censées les expliquer. Il s'agit par exemple de fournir des informations qui permettront d'expliquer, pour une entrée donnée, la prédiction d'un modèle d'apprentissage profond entraîné, e.g. pourquoi ce modèle de type Vision Transformer fournit ce résultat de classification pour une image donnée ? Des techniques de XAI (eXplainable AI) sont proposées et étudiées par la littérature. Elles permettent d'évaluer le comportement global des modèles entraînés, voire d'expliquer des prédictions spécifiques en fonction des entrées traitées (méthodes d'interprétabilité locale). Ces techniques de XAI se basent sur des postulats quant à l'appréciation des modes de traitements des modèles prédictifs, et de l'importance donnée aux informations qui leur sont fournies.

Quelle est l'incidence des techniques de XAI sur la collaboration Homme-Machine ? La mission de postdoctorat proposée dans cette étude vise à fournir des éléments de réponse à cette question. Ainsi, dans de nombreux cas, les modèles d'IA sont des systèmes complexes qui traitent d'énormes quantités de données et effectuent des calculs difficiles à interpréter intuitivement. L'objectif de l'explicabilité est de fournir aux utilisateurs des informations sur la logique de prise de décision du modèle de manière à ce qu'ils puissent comprendre et faire confiance aux résultats de l'IA. Si l'intelligibilité des conclusions d'une IA peut modifier la confiance accordée à cette IA, la collaboration Homme-Machine peut-elle évoluer vers une relation de médiation, de délégation ou de substitution ? Ce sont les questions que nous nous proposons d'aborder dans le cadre de cette étude.

Description de l'emploi

Vous contribuez à :

- ▶ L'évaluation des variations des résultats produits par les méthodes XAI dans des contextes d'étude spécifiques (par exemple, des tâches de classification d'images et des techniques XAI basées sur l'interprétabilité locale à l'aide de méthodes d'attribution).
- ▶ L'évaluation de l'impact des méthodes XAI sur la collaboration homme-machine dans des contextes simplifiés de prise de décision :
 - o évaluation de la performance de l'opérateur humain dans l'exécution d'une tâche, dans différents contextes : seul, avec l'aide d'un modèle prédictif pour lequel les décisions seront expliquées/non expliquées, à l'aide d'une technique XAI.



o évaluation de la collaboration homme-machine : délégation, substitution ou médiation.

o évaluation des biais potentiels induits par les techniques d'IAO.

Vous participez à la définition :

- ▶ Des contextes d'étude (par exemple les jeux, la classification d'images) et des protocoles de test à prendre en compte.
- ▶ A la sélection et à l'implémentation de modèles prédictifs et de techniques XAI.
- ▶ A la mise en place des outils nécessaires à la réalisation des expériences couvertes par les protocoles d'étude, par exemple le développement de jeux simples, d'interfaces décisionnelles.
- ▶ A la mise en œuvre des protocoles susmentionnés sur des cohortes d'opérateurs humains.
- ▶ L'évaluation et la valorisation des résultats obtenus.

En fonction de votre profil, certains aspects sont plus approfondis que d'autres, et des aspects complémentaires sont envisagés concernant l'identification de techniques XAI susceptibles de répondre aux limitations des techniques existantes identifiées dans la littérature ou par les tests réalisés. Nous restons ouverts aux propositions de contributions des candidats en fonction de leurs centres d'intérêt.

Profil recherché et critères généraux d'évaluation

Compétences, connaissances et expériences requises :

- ▶ Modèles d'apprentissage profond, et leur mise en œuvre via PyTorch (capacité à entraîner et à affiner des modèles pré-entraînés sur des ensembles de données spécifiques sur des ressources informatiques GPU dédiées), évaluation des modèles entraînés selon des protocoles standard.
- ▶ Techniques XAI ; connaissance des principales méthodes XAI (par exemple les méthodes locales et d'attribution) et des outils dans le domaine. Les compétences peuvent être améliorées au cours de la mission, mais la connaissance de ces aspects est souhaitable.

Compétences, connaissances et expériences appréciées :

- ▶ Une partie importante de la mission est l'évaluation des collaborations homme-machine, c'est-à-dire l'évaluation de l'impact des modèles d'IA et de XIA sur la prise de décision humaine ; une première expérience de travail avec des cohortes humaines serait un plus.

Niveau de formation et/ou expérience minimums requis :

- ▶ Doctorat en informatique sur une thématique en lien avec l'apprentissage automatique et/ou profond

Candidature



Conditions administratives de candidature

Le poste proposé par l'IMT Mines Alès est un contrat à durée déterminée de 18 mois, à temps plein, contrat de droit public relevant des dispositions du cadre de gestion de l'Institut Mines-Télécom, métier P, Post Doctorant, catégorie II. Le projet sera mené en étroite collaboration avec IMT Business School (BS), en particulier avec Nicolas Soulié. Le poste sera localisé a priori à IMT Mines Ales, néanmoins une localisation à IMT BS (Evry, Ile-de-France) pourrait être examinée.

Salaire : à définir selon le profil et expérience.



Modalités de candidature

Les candidatures (CV et lettre de motivation) sont à adresser exclusivement à :

<https://institutminestelecom.recruitee.com/o/post-doctorant-ou-post-doctorante-en-apprentissage-automatique>



IMT Mines Alès
École Mines-Télécom



Planning du recrutement

Date limite de clôture des candidatures : 12/05/2024

Date presentie indicative du jury : 2ème quinzaine de mai 2024

Date de prise de fonction souhaitée : 01/07/2024



Personnes à contacter

► Sur le contenu du poste :

Sébastien HARISPE, Enseignant-chercheur

✉ : sebastien.harispe@mines-ales.fr

Nicolas SOULIE, Enseignant-chercheur

✉ : nicolas.soulie@imt-bs.eu

► Sur les aspects administratifs :

Géraldine BRUNEL, responsable du service des relations humaines

✉ : geraldine.brunel@mines-ales.fr