

# Appel Stages Master 2 2024

## Introduction

L'institut DATAIA est l'institut de convergence français dédié aux sciences des données, intelligence artificielle et leurs impacts sociétaux. Il est au carrefour des technologies big data et de l'intelligence artificielle avec les autres disciplines, notamment les sciences humaines et sociales, pour une intelligence artificielle au service de l'humain. Sélectionné par l'Agence Nationale de la Recherche (ANR) dans le cadre des « Programmes d'Investissements d'Avenir », l'Institut DATAIA a pour vocation de regrouper et de structurer des expertises pluridisciplinaires pour mieux répondre à des enjeux majeurs des sciences des données, de l'intelligence artificielle et de leurs applications.

L'institut soutient la recherche pluridisciplinaire en Science des données au sens large — incluant l'informatique, les mathématiques, les sciences physiques, les sciences de la vie et les sciences humaines et sociales— à travers un appel annuel dédié aux stages de Master 2.

## Pré-requis

Le porteur du projet doit appartenir à un laboratoire d'un établissement partenaire DATAIA listé en annexe 2 de l'appel. La proposition doit rentrer dans le cadre des problématiques de DATAIA décrites en annexe 1. Les projets impliquant plusieurs laboratoires d'établissements partenaires DATAIA différents seront fortement appréciés.

## Dates importantes

|          |                         |
|----------|-------------------------|
| Clôture  | 10 novembre 2023 midi   |
| Décision | 08 décembre 2023        |
| Contact  | contact-dataia@inria.fr |

## Procédure de soumission

Le dossier soumis doit inclure :

### Les informations obligatoires :

- Titre du projet ;
- Noms des encadrants et leurs coordonnées ;
- Noms des laboratoires d'accueil et/ou des équipes de recherche.

### La description du projet (2 pages) :

Les partenaires décrivent l'objectif du stage, son lien avec DATAIA, et listent les collaborations antérieures des partenaires (le cas échéant). Il n'est pas indispensable d'avoir un candidat pour la soumission du projet.

### L'annexe financière indiquant :

- Montant de l'aide demandé ainsi que la durée du stage ;
- Dates de début et fin de stage envisagées ;
- Nom de l'établissement qui reçoit et gère les fonds versés ;
- Les noms et coordonnées des contacts administratifs et financiers en charge du suivi du projet

Un modèle de réponse est disponible à l'adresse suivante :

<https://dataia.eu/appel-projets/appel-stages>

### Le dossier doit être envoyé par le laboratoire d'accueil à :

[submission-dataia@inria.fr](mailto:submission-dataia@inria.fr)

### Soutien :

Si le projet est accepté, DATAIA financera l'étudiant pour 4 à 6 mois au niveau de gratification habituel (environ 650 euros par mois).

Pour information :

- L'étudiant doit être présent plus de 308 heures pendant la même année d'enseignement pour bénéficier d'une gratification ;
- Les organismes publics ne peuvent pas verser de gratification de stage supérieure au montant minimum légal sous peine de requalification de la convention de stage en contrat de travail ;
- Le montant légal minimum en 2023 pour une gratification de stage est de 4,05 € net de l'heure et aucune charge à régler ;
- Le plafond horaire de 2024 ne sera connu qu'à partir de l'arrêté au Journal officiel en novembre ou décembre 2023.

### Engagements :

Le rapport de stage devra être adressé à DATAIA à l'issue du stage. Une présentation brève pourra être demandée lors des journées scientifiques annuelles de DATAIA

Toute communication (publication, présentation) devra obligatoirement mentionner le financement DATAIA :

*This research was supported by DATAIA convergence institute as part of the "Programme d'Investissement d'Avenir", (ANR- 17-CONV-0003) operated by [Partner] XXX*

## **Annexe 1 : Axes de recherche de DATAIA**

Le programme scientifique de chaque proposition devra aborder explicitement au moins un des quatre axes de recherche esquissés ci-dessous, qui sont les quatre défis interdisciplinaires de DATAIA.

### **DES DONNÉES AUX CONNAISSANCES, DES DONNÉES À LA DÉCISION**

La disponibilité croissante de données massives repousse les frontières techniques dans de nombreux champs. D'une part, la nature hétérogène, semi-structurée, incomplète ou incertaine des données remet en cause les modèles statistiques usuels ainsi que les algorithmes dédiés à la décision. D'autre part, la gestion des données soulève de nouvelles contraintes d'opérabilité telles que la sécurité, l'intégrité et la traçabilité.

En outre, produire de la connaissance nécessite de construire des modèles qui fournissent des décisions explicables, statistiquement valides et calculables. L'acceptation des résultats requiert aussi que la confidentialité et la loyauté soient renforcées. En parallèle, de nouveaux développements en optimisation doivent permettre d'améliorer les procédures d'estimation.

#### **Défis :**

- Données hétérogènes, complexes, incomplètes, semi-structurées et/ou incertaines
- Données massives : algorithmique et structuration des données
- Apprentissage en ligne, méthodologie pour les données massives, méthodes efficaces
- Amélioration du stockage, calcul et estimation pour la science des données
- Modélisation des interactions entre agents (humains ou artificiels) par théorie des jeux
- Représentation et algorithmes multi-échelles et multimodaux
- Analyse théorique de méthodes heuristiques (théorie de la complexité, géométrie de l'information, théorie des chaînes de Markov)
- Coévolution Humain-Machine dans les systèmes autonomes : agents conversationnels, voitures, robots sociaux

### **APPRENTISSAGE ET INTELLIGENCE ARTIFICIELLE**

Récemment, la recherche en apprentissage profond a réalisé des avancées spectaculaires en vision par ordinateur et en traitement du langage naturel. Au-delà de l'arrivée des données massives, de l'accroissement de la puissance de calcul et des efforts de conception, les causes de ces progrès, encore mal connues, soulèvent au moins trois questions. Quelle théorie de l'apprentissage permettra d'analyser les architectures profondes ? Comment gérer la compositionnalité de ces architectures et leur capacité à appréhender des objets plus complexes ? Comment ouvrir la boîte noire pour mettre à jour les représentations apprises ?

## Défis

- Apprentissage automatique innovant et IA : sens commun, adaptabilité, généralisation
- Apprentissage profond et apprentissage adversarial
- Apprentissage automatique et hyper-optimisation
- Optimisation pour l'apprentissage (e.g. Améliorations des méthodes de gradient stochastiques, optimisation Bayésienne), optimisation combinatoire
- Lien apprentissage-modélisation, intégration à priori dans l'apprentissage
- Reproductibilité et apprentissage robuste
- Inférence statistique et validation
- Compositionnalité des architectures profondes.

## TRANSPARENCE, IA RESPONSABLE ET ÉTHIQUE

La confiance numérique s'instaure à partir de l'implantation de méthodologies éthiquement responsables à travers la transparence et la redevabilité des systèmes algorithmiques ; de la régulation de la collecte, de l'usage et du traitement de données personnelles ; du renforcement de la régulation au travers de procédures numériques appropriées.

La confidentialité par conception est une forme de régulation qui inclut la protection des données personnelles dans toutes les étapes de la collecte et du traitement. Le traçage des outils appliqués aux données doit également être développé de manière à faciliter l'explication du modèle pour les experts comme pour les utilisateurs rendant les systèmes algorithmiques auditables. Les principes de confidentialité, bien que faciles à formuler, nécessitent de modifier les infrastructures de stockage et de traitement, avec d'importants impacts législatifs, sociologiques et économiques. Les techniques de transparence des systèmes algorithmiques seront développés en focalisant sur : l'équité, la loyauté et la non-discrimination et la redevabilité-par-construction.

## Défis

- Responsabilité et explicabilité par conception
- Transparence et équité par conception
- Audit des systèmes algorithmiques : non-discrimination, loyauté, biais techniques, neutralité, équité
- Mesure de la confiance et de l'appropriation du numérique
- Progressive user- analyse de donnée progressive (monitoring interactif des systèmes de décisions)
- Responsabilité du traitement de l'information et de la prise de décision : contrôle d'usage des données et vérification des faits
- Découverte causale, traçabilité des inférences à partir des données sources, interprétabilité des architectures profondes

## PROTECTION, RÉGULATION ET ÉCONOMIE DE LA DONNÉE

Les entreprises impliquées dans l'économie de la donnée ont continuellement besoin de repenser leur structuration : elles doivent adopter une organisation orientée-projet avec des changements rapides dans l'allocation de ressources. L'économie de la donnée soulève aussi des problèmes de concentration et de monopole.

Un petit nombre d'entreprises (GAFAM) détiennent la plupart des données. Cette

concentration de marché peut conduire à une compétition déloyale dont l'innovation dans les petites et moyennes entreprises risque de pâtir. Les citoyens attendent des gouvernements qu'ils interviennent dans l'économie digitale pour éviter trop de concentration et de monopole. Les gouvernements doivent empêcher la fuite des informations pour préserver la souveraineté des états et le respect des réglementations.

### Défis

- Respect de la vie privée par conception GDPR
- Apprentissage respectueux de la vie privée (*differential privacy* )
- Développement de méthodologies éthiquement responsables, et technologies pour réguler la collecte, l'usage et le traitement des données personnelles, et l'exploitation des connaissances tirées de ces données
- Sécurité informatique des chaînes de traitement de données
- Sécurité/crypto : block-chain et tiers de confiance

Les projets pourront aborder des sujets plus spécifiques à des domaines d'application. Une liste de ces domaines est donnée ci-dessous à titre indicatif. Toutefois, il est demandé que les projets abordent des questions de science des données d'intérêt général, c'est-à-dire non restreintes spécifiquement à un domaine d'application.

- Energie : optimisation de la gestion (production et distribution), et régulation
- Santé, nutrition personnalisée et bien-être
- Urbanisation Mobilité (véhicule connecté et autonome, villes intelligentes. . .)
- Analyse pour la finance les assurances
- Internet des personnes et des objets
- E-Sciences.

## **Annexe 2 : Liste des partenaires de DATAIA éligibles à l'appel**

- Agro ParisTech
- CentraleSupélec
- CEA
- CNRS
- ENS Paris-Saclay
- IFP- Énergies Nouvelles
- Inria
- Institut Mines-Télécom Business School
- Inrae
- ONERA
- Université d'Evry Val d'Essonne
- Université Paris-Saclay
- Université Versailles St-Quentin-en-Yvelines