

# Appel à projets

## Institut Convergence DATAIA

### Contexte général de l'appel

L'Institut DATAIA est l'institut de convergence français spécialisé en sciences des données, intelligence artificielle et société, ou comment les technologies big data et l'intelligence artificielle se croisent avec 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 de grande ampleur et de forte visibilité pour mieux répondre à des enjeux majeurs des sciences des données, de l'intelligence artificielle et de leurs applications grâce au décloisonnement entre les mathématiques, les sciences informatiques et les sciences juridiques, économiques et sociales.

Il mobilise à son lancement plus de 130 chercheurs et enseignants-chercheurs de dix-sept établissements (voir annexe 2 ci-dessous), ainsi qu'un programme d'affiliation industrielle permettant la collaboration étroite avec les entreprises.

L'Institut DATAIA soutient la recherche en Data Science à travers un "Appel à projets recherche" annuel.

### Périmètres scientifiques de l'appel et objectifs

Les réponses à cet appel devront montrer en quoi elles concourent aux objectifs de l'Institut DATAIA et notamment à la structuration du domaine de la science des données au sein du Campus Paris-Saclay. En conséquence, toutes les propositions doivent s'inscrire dans au moins un des quatre défis principaux de DATAIA et peuvent se rapporter à un domaine applicatif (voir la liste des « Axes de travail » en annexe 1). Elles devront s'appuyer sur la collaboration d'au moins deux personnes de deux établissements partenaires de l'institut DATAIA (voir la liste en annexe 2) n'appartenant pas au même laboratoire ni au même établissement hébergeur.

Cet appel 2019 visera au soutien des actions de DATAIA au travers de financement de trois projets : chaque projet de recherche inclut le financement d'une thèse et d'un CDD de 2 ans (post-doc ou ingénieur) ou de deux thèses et éventuellement des frais de fonctionnement. La demande budgétaire par projet ne pourra pas excéder 220k€ ;

*NB : des appels spécifiques et permanents sont dédiés à l'organisation de workshops et à l'invitation de professeurs, au fil de l'eau.*

### Périmètre partenarial de l'appel

Cet appel est réservé aux équipes relevant du périmètre de l'institut de convergence DATAIA. Seuls les laboratoires des établissements partenaires figurant dans la convention de financement pourront être porteurs et gestionnaires des crédits alloués (cf. liste en annexe 2). Les consortiums pourront compter des partenaires hors des membres fondateurs de DATAIA mais ceux-ci ne seront pas financés.

## Modalités de soumission des propositions

Chaque dossier comportera une partie de description scientifique du projet et une annexe financière. Les dossiers soumis hors délai ou incomplets ne seront pas examinés.

Format du dossier de soumission :

Le(s) porteur(s) de projet soumet(tent) un document comportant les informations suivantes :

- ACRONYME et titre du projet
- Nom, laboratoire et coordonnées des porteurs
- Les laboratoires (ou équipes) de l'institut de convergence qui bénéficieront des financements, le contact scientifique et le contact administratif au sein de chaque laboratoire (équipe) du partenaire bénéficiaire
- Durée et dates envisagées du projet
- Projet scientifique (5 pages max) :

Le projet décrira le contexte du projet et l'état de l'art, les objectifs du travail de recherche, son caractère innovant et les verrous, sa valeur ajoutée, les critères de succès, les pistes de travail envisagées, sa portée applicative, ainsi que l'adéquation avec les objectifs et axes de travail de l'institut de convergence.

Pour l'évaluation des projets, les critères suivants seront considérés : l'excellence scientifique, la synergie entre les partenaires et l'impact applicatif potentiel lié au sujet traité.

Tous les projets doivent être soumis sur le site suivant : <https://cmt3.research.microsoft.com/DATAIA2019>

Tous les projets sont à soumettre avant le **1<sup>er</sup> mars 2019 à 12h** (heure de Paris).

Cet appel à projets est disponible sur le site <http://dataia.eu>

Vous pouvez poser vos questions sur cet appel à :

- Viviane Hoang ([viviane.hoang@inria.fr](mailto:viviane.hoang@inria.fr))
- Frédéric Pascal ([frederic.pascal@centralesupelec.fr](mailto:frederic.pascal@centralesupelec.fr))

Résultats de l'appel : **26 avril 2019**

Mise à disposition des crédits alloués à partir de mai 2019 et début de contractualisation entre l'Université Paris-Saclay avec les bénéficiaires.

## Modalité de sélection des projets

L'éligibilité des projets sera traitée par les membres du bureau. Une pré-sélection des projets éligibles sera effectuée par le comité des programmes, qui pourra avoir recours à des rapporteurs extérieurs. Les porteurs des propositions pré-sélectionnées passeront une audition devant le comité des programmes qui délibérera pour décider des projets retenus. Les décisions finales seront publiées le **26 avril 2019**.

## Validité des financements DATAIA

Les projets de recherche sont financés pour une durée de 3 ans, et doivent être menés à terme le 31/12/2022. Chaque thèse financée devra impérativement démarrer en 2019.

## Engagement en cas d'acceptation

En cas de financement du projet, il conviendra de communiquer à l'institut de convergence le plus rapidement possible le CV du/des candidat(s) retenu(s) (doctorant, post-doctorant ou ingénieur). La validation des candidats en thèse est conditionnée par l'approbation de l'école doctorale.

Le porteur de projet s'engage à utiliser les crédits conformément à la demande validée

par le comité des programmes de DATAIA, à rendre compte annuellement de l'activité scientifique et de l'utilisation des crédits. En fin de projet, il remettra un rapport décrivant les réalisations et listant les publications effectuées dans le cadre du projet.

Les bénéficiaires du soutien de l'institut de convergence doivent faire apparaître dans toutes les communications scientifiques (ou non) associées aux activités financées la mention de DATAIA et des programmes d'investissements d'Avenir (PIA) en insérant la mention : « 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. »

Si l'institut de convergence juge que les crédits n'ont pas été utilisés conformément aux objectifs annoncés, il pourra être demandé un remboursement. De même, les crédits non utilisés devront être remboursés par l'établissement partenaire.

## **Annexe 1**

### **Axes de travail**

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 : structuration de la donnée pour pouvoir l'exploiter
- 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-échelle 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 d' priori dans l'apprentissage
- Reproductibilité et apprentissage robuste
- Inférence statistique et validation
- Compositionnalité des architectures profondes Apprentissage profond et apprentissage adversarial ; Apprentissage automatique et hyper-optimisation ; Optimisation pour l'apprentissage, e.g., améliorations des méthodes de gradient stochastique, optimisation bayésienne, optimisation combinatoire ; Lien apprentissage-modélisation, intégration d'a 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--analytics » analyse de donnée progressive (monitoring interactif des systèmes de décisions : dataviz dashboards, IHM)
- 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.

Challenges :

- « -by-design » respect de la vie privée par conception, GDPR
- Apprentissage respectueux de la vie privée( « *differential privacy* »)
- Développement de méthodologies éthiquement responsables, et de technologies pour réguler la collection, 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.

## **Annexe2**

### **Liste des partenaires de DATAIA**

Agro ParisTech

CentraleSupélec

CEA

CNRS

Ecole polytechnique

ENS Paris-Saclay

ENSAE

ENSTA ParisTech

HEC Paris

IFP-EN

Inria

Institut Mines-Telecom (Télécom ParisTech, Télécom Sud Paris, Institut Mines-Télécom Business School)

Inra

Université d'Evry Val d'Essonne.

Université Paris-Saclay

Université Paris-Sud

Université Versailles St-Quentin-en-Yvelines