APPEL A MANIFESTATION D'INTERET

Programme « confiance.ai »

Décembre 2020

1 Le programme « confiance.ai »

L'intelligence artificielle, aujourd'hui critique pour la compétitivité de l'industrie française, demande encore de nombreuses recherches et innovations afin d'atteindre tout son potentiel. En particulier, **l'intégration et l'utilisation sûre des technologies d'IA** sont essentielles pour soutenir l'ingénierie, le développement et la diffusion de produits et services innovants.

« L'industrialisation de l'intelligence artificielle » pour les systèmes critiques est un des objectifs majeurs du Grand Défi « Sécuriser, certifier et fiabiliser les systèmes fondés sur l'intelligence artificielle »[1], un des grands défis portés par le Conseil de l'Innovation, qui bénéficie d'un financement global de 30M€ sur une durée de quatre ans, complété par les apports des industriels et des établissements.

Confiance.ai, premier des trois piliers du Grand Défi, vise à apporter un environnement en support à la conception, à la validation et au test pour renforcer la confiance, l'explicabilité et avancer vers la certification des systèmes d'IA. Jusqu'à 20% du budget du programme confiance.ai est destiné aux contributions académiques.

En répondant à l'Appel à Manifestations d'Intérêt de confiance.ai, vous avez aujourd'hui la possibilité de rejoindre un collectif d'industriels et d'académiques majeurs[2] et motivés pour participer à une des actions prioritaires de la stratégie nationale d'intelligence artificielle, et ce en confrontant et faisant progresser vos méthodes, algorithmes et outils sur des cas d'utilisation critiques pour de nombreux secteurs d'application – notamment ceux de l'énergie, du transport et de la défense/sécurité.

1 https://www.gouvernement.fr/grand-defi-securiser-certifier-et-fiabiliser-les-systemes-fondes-sur-l-intelligence-artificielle

2 Industriels : Air Liquide, Airbus, Atos, EDF, Renault, Naval Group, Safran, SopraSteria, Thales, et Valeo. Académiques : CEA, Inria ; IRT Saint Exupéry et IRT SystemX

2 Objectif de l'AMI et thématiques scientifiques

L'objectif de cet appel est de solliciter les laboratoires académiques pour contribuer à la résolution des verrous scientifiques liés au programme confiance.ai. Ces verrous se rapportent aux trois thématiques suivantes, plus de détail étant fourni en annexe :

- confiance et ingénierie système à composants IA ;
- confiance et données d'apprentissage ;
- confiance et interaction humaine.

Le programme est organisé en cinq projets de développement d'outils et méthodes pour la confiance dans les systèmes à base d'IA, plus deux projets de support à l'ensemble. Vos propositions sont les bienvenues sur la totalité du programme.

Projets support :

1) Intégration et cas d'utilisation

Ce projet réalisera le socle fédérateur de confiance.ai pour les briques et processus méthodologiques réalisés par les autres projets, afin de permettre une accélération de la mise en œuvre de composants IA dans les systèmes critiques ; il établira les cas d'usages industriels qui serviront à évaluer la performance des méthodes et outils développés.

2) Processus, méthodologies, outils

Ce projet revisite les ingénieries classiques (algorithmique, logicielle et système) au regard des exigences de l'IA de confiance : qualification, homologation, certification, couverture du domaine opérationnel, biais, sûreté et réponse aux attaques malveillantes, interaction humain-système, déploiement à grande échelle, etc.

Projets de développement :

3) Caractérisation et qualification de l'IA de confiance

Ce projet vise à caractériser les logiciels de traitement des données et à proposer des solutions pour la détection en ligne de données anormales (hors du domaine d'apprentissage) ; caractérisation de la robustesse d'une IA basé données ; évaluation de la qualité d'un composant IA basé données ou de connaissances ; méthodes et métrologie de l'explicabilité ; outils et solutions de supervision des IA.

4) Conception de systèmes d'IA de confiance (algorithmes et outils)

Ce projet fournira des méthodes permettant de garantir la confiance a priori par construction lors du processus de développement : méthodes d'apprentissage garantissant la confiance par construction ; méthodes de conception d'IA à base de connaissances garantissant la confiance par construction ; outils supportant ces méthodes ;

5) Ingénierie de la connaissance et des données pour l'IA de confiance

Ce projet fournira des outils permettant de gérer la donnée, l'information et la connaissance au long du cycle de vie d'une application : formalisation des exigences, annotation, caractérisation, visualisation des données et connaissances ; augmentation de jeux de données ; gestion de la frugalité ; évaluation de la couverture du domaine d'opération etc.

6) Intégration, Vérification, Validation, Qualification (IVVQ) de l'IA de confiance – vers la certification

Ce projet revisite les processus d'IVVQ au regard des spécificités de l'IA : spécification des exigences, stratégies d'IVVQ, gestion des risques, qualification incrémentale et évolutive, le tout selon des niveaux de criticité dépendant des usages et domaines visés.

7) IA de confiance embarquée

Ce projet traite les difficultés particulières posées par l'intégration de composants IA dans les systèmes embarqués au regard de la confiance : contraintes physiques (taille, poids, énergie..), performances temps réel, sûreté de fonctionnement et disponibilité, réglementations

spécifiques, etc.

3 Modalités des réponses

3.1 Format des contributions

Les réponses à l'AMI ne devront pas excéder 3 pages et comprendront les éléments suivants :

- Présentation du laboratoire, de l'équipe, des chercheurs et chercheuses impliqués
- Contenu scientifique de la proposition, outils et méthodes existants, bibliographie
- Modalité(s) de collaboration souhaitée(s) parmi les quatre ci-dessous.
- Estimation du budget et de la durée nécessaire du projet.

  Un canevas de réponse est fourni.

• Contribution scientifique ponctuelle

- *Objectif* : Il s'agit d'une étude ponctuelle d'état de l'art, d'analyse ou de mise en œuvre de travaux de recherche existants
- *Outil* : contrat de type prestation avec un laboratoire académique (à réaliser en interne du laboratoire, par un chercheur, une chercheuse, un ou une postdoc, un ou une ingénieur(e) de recherche …)
- Le financement des contributions retenues se fera à 100% des coûts marginaux ou 50% du coût complet selon la nature du laboratoire.

• Montée en TRL de travaux scientifiques

- *Objectif* : Il s'agit de mener des travaux de recherche plus appliqués partant d'un état de l'art de travaux amont
- *Outil* : postdoc embauché(e) par l'IRT SystemX avec contrat d'encadrement pour le laboratoire
- Le financement des contributions retenues se fera à 100 % du temps d'encadrement, le ou la postdoc étant directement salarié(e) de SystemX.

• Résolution d'un verrou scientifique amont

- *Objectif* : Il s'agit d'un travail de recherche amont, typiquement sous la forme de financement d'un doctorat ou un postdoc, apportant des éléments de réponses aux verrous identifiés par le projet
- *Outil* : thèse/postdoc embauché(e) par l'IRT SystemX, avec contrat d'encadrement pour le laboratoire ou mise à disposition (MAD) auprès de l'IRT d'un encadrant académique de la thèse/postdoc pour sa supervision.
- Le financement des contributions retenues se fera à 100% du temps d'encadrement, le chercheur ou la chercheuse étant directement salarié(e) de SystemX.

• Autre mode de contribution

- *Préciser objectif et outil proposé.*

3.2 Processus de sélection et démarrage des travaux

- Ouverture de l'appel : **10/12/2020**
- Réunion d'information (en ligne sur le lien) : **5/1/2021 de 14h à 16h** à destination de tous les laboratoires, chercheuses et chercheurs intéressés.
- Date limite de réception des réponses : **25/1/2021 à 12h.**

- Contact pour l'élaboration des réponses et envoi des réponses : *bertrand.braunschweig@irt-systemx.fr*

Un comité de pré-sélection composé de représentants des partenaires du programme, animé par le coordinateur scientifique, examinera les réponses par rapport aux critères suivants :

- Qualité de l'équipe ;
- Intérêt scientifique de la proposition et pertinence par rapport aux objectifs du programme ;
- Budget demandé.

Le programme confiance.ai sélectionnera les répondants sur la base de l'avis du comité de pré-sélection.

A l'issue du dépouillement des réponses à l'AMI, des ateliers seront organisés au **1er trimestre 2021** rassemblant partenaires industriels et répondants sélectionnés pour affiner les réponses sur les cas d'utilisation qui seront traités dans le programme et définir les sujets de thèse devant démarrer à la rentrée 2021.

Les travaux proprement dits pourront démarrer à partir du **troisième trimestre 2021**.

3.3 Intérêt à suivre le programme, sans contribution

Le programme organisera périodiquement des sessions d'information, journées scientifiques pour diffuser certains résultats. Si vous souhaitez être informés de ces évènements, répondez simplement à cet appel sans proposer de contribution, avec une simple présentation de l'équipe, de la ou des personnes concernées, et les informations de contact nécessaires.

4

5 Annexe : thématiques identifiées

Certains verrous et problématiques scientifiques et technologiques ont déjà été identifiés par les partenaires au cours du montage du programme, regroupés en trois grandes thématiques décrites ci-dessous. Ils sont donnés **de manière non exhaustive**, toute contribution pertinente sur le thème de la confiance envers les systèmes d'IA étant a priori bienvenue.

5.1 Thématique « Confiance et ingénierie système à composants IA »

• Construire des composants IA à confiance maitrisée

- Modèles d'apprentissage intégrant de la connaissance (modèle physique, sémantique) ;
- Garanties théoriques de robustesse des réseaux de neurones sur les mauvais usages ;
- Composants intégrant une auto-surveillance de détection de sortie de leur zone de fonctionnement nominal.

• Qualifier des composants et systèmes à base d'IA

- Métrologie, caractérisation et quantification du risque d'un composant / système intégrant de l'IA
- Prise en compte du caractère évolutif et incertain du risque d'un composant / système intégrant de l'IA
- Qualification de solutions utilisant des réseaux pré-entrainés (transfer learning)

• Embarquabilité de l'IA de confiance

- Garantir la conformité du modèle embarqué au modèle de développement
- Estimation des performances (temps d'exécution, énergie) d'un composant d'IA à partir de ses paramètres, et de l'architecture hardware cible
- Identifier l'impact de faute matérielle sur le comportement des composants ML
- Exploitation des propriétés de régularité des algorithmes de ML sur les architectures hardware hétérogène

5.2 Thématique « Confiance et données d'apprentissage »

• Construire les données/connaissances pour augmenter la confiance dans l'apprentissage

- Intégrer de la connaissance physique par génération de données d'apprentissage simulées (EDP)
- Intégrer de la connaissance comportementale par génération de données d'apprentissage simulées (SMA)
- Maitriser la gestion du cycle de la donnée en IA
- Maitriser l'apprentissage actif, l'automatisation de l'annotation et la couverture associée construite

• Qualifier les données/connaissances pour l'apprentissage

- Mesurer la qualité et la représentativité des données/connaissances d'un domaine opérationnel
- Evaluer la qualité du processus d'exploration, enrichissement, annotation et préparation des données
- Identifier les biais dans les données / connaissances conduisant à de mauvaises décisions
- Définir des méthodes et langages d'élaboration des spécifications fonctionnelles sous la forme de jeux de données
- Définir des métriques d'évaluation de la couverture d'un domaine opérationnel pour un jeux de données et/ou d'information
- Construire des fonctions de représentation de la variabilité du domaine opérationnel autour des données disponibles pour le décrire
- Détecter les écarts de représentativité des données d'opération avec les données d'apprentissage

5.3 Thématique « Confiance et interaction humaine »

• Interaction générant de la confiance entre utilisateur - système à base d'IA

- Méthodologie de prise en compte de la confiance dans la relation humain système*
- Méthodologie de prise en compte de l'acceptabilité par les utilisateurs

• Interaction concepteur/certificateur - système à base d'IA

- Approches d'explication de modèles prédictifs de type boite noire (suivant typologie d'utilisateur)
- Métrologie et caractérisation d'un module d'explication
- Explicabilité des approches hybrides connaissance/apprentissage/optimisation

Les membres du collectif :