COMMUNIQUE DE PRESSE

CONFIANCE.AI DAYS

LE COLLECTIF CONFIANCE.AI DEVOILE LES AVANCEES SCIENTIFIQUES ET TECHNOLOGIQUES DE SON AMBITIEUX PROGRAMME DEDIE A L'IA DE CONFIANCE DANS LES SYSTEMES CRITIQUES

Initiative lancée dans le cadre de la 1ère phase de la Stratégie Nationale en IA et financée par France 2030, le programme Confiance.ai relève le défi ambitieux de développer un environnement méthodologique et technologique au service de l'intégration de l'IA de confiance dans les systèmes critiques. Après 20 mois de collaboration étroite au sein d'un écosystème riche de près de 50 partenaires industriels et académiques, une première version de l'environnement de confiance est d'ores et déjà déployée au sein des ingénieries des partenaires et une deuxième version contenant des avancées majeures est en cours de développement.

Paris, le 04 octobre 2022

A l'occasion des Confiance.ai Days, les membres fondateurs du collectif Confiance.ai (Air Liquide, Airbus, Atos, Naval Group, Renault, Safran, Sopra Steria, Thales, Valeo, ainsi que le CEA, Inria, l'IRT Saint Exupéry et l'IRT SystemX) présentent les avancées scientifiques et technologiques du programme, les retours des premiers déploiements chez les partenaires et les perspectives à venir.

Pilier technologique du Grand Défi « Sécuriser, fiabiliser et certifier des systèmes fondés sur l'intelligence artificielle », le programme Confiance.ai vise à répondre à la problématique de l'intégration d'une IA sûre, fiable et sécurisée dans les systèmes industriels critiques (véhicules autonomes, contrôle industriel en ligne, systèmes d'aide à la décision, …). Pour ce faire, les partenaires conjuguent, sur une durée de 4 ans, leurs expertises en IA, ingénierie et sûreté de fonctionnement pour lever les verrous associés à l'industrialisation de l'IA et doter les acteurs industriels européens d'un environnement de confiance (suite d'outils logiciels et de nouvelles méthodes) souverain, ouvert, interopérable et pérenne.

Un programme contribuant à la mise en œuvre opérationnelle du futur AI-Act

Le déploiement à très grande échelle de systèmes industriels intégrant une IA de confiance est un enjeu clé de compétitivité industrielle, économique, mais aussi de souveraineté. L'Union Européenne, avec son projet de règlement sur l'IA intitulé « Artificial Intelligence Act » souhaite encadrer l'intelligence artificielle et ses usages, pour la rendre digne de confiance, éthique, durable, inclusive et centrée sur l'humain. Elle vise à instaurer un cadre réglementaire et juridique qui régira les technologies d'IA conçues au sein des pays membres mais aussi celles des opérateurs traitant avec eux, en les classant en 4 catégories selon leur niveau de risque. Le programme Confiance.ai contribuera à la mise en œuvre opérationnelle de ce futur règlement par les industriels, en leur proposant un environnement technique qui garantira un haut niveau de confiance dans les produits et services à base d'IA.

Un collectif composé de 50 partenaires

L'une des spécificités de ce programme est sa démarche intégrative et ouverte. Il fédère un écosystème de partenaires industriels, académiques et de laboratoires de recherche autour de

l'ambition de faire de la France un leader de l'IA de confiance. Ainsi, autour des 13 fondateurs s'est rapidement constitué un collectif de près de 50 partenaires industriels et académiques. Ont notamment rejoint fin 2021 le programme 12 start-up et PME lauréates d'un Appel à Manifestation d'Intérêt (AMI) ; elles viennent enrichir les travaux du programme avec leurs technologies de simulation, d'interactions homme-machine, de test et d'explicabilité. Un programme doctoral de 8 thèses et 4 postdocs a également été mis en place fin 2021, suite à un AMI académique. Sont désormais partenaires du projet les laboratoires IRIT Toulouse, Onera, Inria, Cristal CNRS, Lamih – Lille, LIP6 – Sorbonne Université, IMT – Toulouse, U2IS-ENSTA, LITIS – INSA Rouen, CRIL – Université Artois, au travers d'un AMI sur les Sciences Humaines et Sociales lancé en 2022. D'autres partenariats stratégiques ont été initiés comme celui avec ANITI en septembre 2022 pour maturer les connaissances produites par ANITI autour de l'IA certifiable et hybride au contact de cas d'usages industriels.

2021 : Une première version de l'environnement de confiance

Concrètement, le programme adresse les grandes thématiques autour desquelles s'articule l'IA de confiance : méthodes et guidelines pour la conception de l'IA de confiance, caractérisation et évaluation de systèmes à base d'IA de confiance, conception de modèles à base d'IA de confiance, ingénierie de la donnée et de la connaissance, certification et IVVQ, et embarquabilité.

52 interviews très riches (140 questions) ont été réalisés auprès des partenaires industriels du programme et ont permis de bien comprendre leurs attentes et ambitions. Il en ressort que si l'intérêt porté sur l'IA et ses promesses est évident, son intégration effective est beaucoup plus timide. Les préoccupations exprimées désignent avant tout l'absence d'un cadre de conception outillé, à la fois fiable et efficace, sur lequel il est possible de s'appuyer et de se référer.

Ont été initiés en parallèle une **approche top-down** (définition et spécification très détaillée des besoins opérationnels du programme et des capacités attendues de l'environnement de confiance) et une **démarche bottom-up** (choix et test de la pertinence de composants, librairies, algorithmes modèles, etc. existants). Cette approche a permis de constituer une **vingtaine d'états de l'art scientifiques conséquents** portant sur les différentes thématiques adressées par l'IA de confiance. Le monitoring, l'ingénierie de la donnée et de la connaissance, l'intelligence artificielle symbolique ou la caractérisation de la notion de confiance sont autant de sujets qui ont fait l'objet d'états de l'art dédiés, offrant une vision d'ensemble complète sur les aspects tant scientifiques que technologiques.

Apportés par les partenaires, **11 premiers cas d'usage** issus de problématiques opérationnelles réelles offrent aux équipes un référentiel concret de contraintes, de modèles, de données et d'objectifs sur lesquels ils peuvent appuyer leurs travaux, tester les différents composants technologiques et méthodologiques identifiés afin d'en valider ou non leur pertinence dans le contexte de l'IA de confiance.

Une **première version de l'environnement de confiance** a été livrée fin 2021. Elle comporte un environnement de développement adossé à une chaine MLOps (chaine de traitement et de déploiement automatique des modèles de machine learning). Mise à disposition des équipes, elle sert d'environnement de travail dans le cadre du programme.

Début 2022 : Un déploiement réussi au sein des ingénieries des partenaires

Au printemps 2022, la première version de l'environnement de confiance a été mise à disposition des industriels. Certains d'entre eux l'ont d'ores et déjà redéployé, ce qui permet d'avoir des premiers retours opérationnels.
C'est le cas de **Safran** qui a ainsi déployé la chaîne outillée de l'environnement de confiance, avec l'ensemble des services opérationnels tels que proposés à l'intérieur du programme. Les premiers retours sont très positifs et il est désormais question d'étendre l'expérimentation de l'environnement au travers sa mise en application sur un cas d'usage interne.

« Nous avons pu aisément adapter la procédure d'installation de l'environnement de confiance avec le support de Confiance.ai. Notre équipe a installé cet environnement sur un de nos serveurs de calcul dans

une logique « On Premise ». Cette opération est stratégique pour nous du fait du caractère sensible de nos activités car nous avons désormais l'opportunité d'appliquer les briques de cet environnement à nos uses cases internes en nous affranchissant du recours à un cloud public. Nous avons prévu d'évaluer dans les prochains mois l'interopérabilité des outils de MLOps avec les outils d'explicabilité et de robustesse développés par Confiance.ai dans le cadre de nos uses cases. Nous attendons avec impatience la version 2 avec les nouveautés annoncées, en particulier autour de la data », explique Jacques Yelloz, ingénieur en chef dans le domaine de l'IA chez Safran Electronics & Defense.

Du côté de **Sopra Steria,** c'est la possibilité de redéployer individuellement les actifs du programme qui représente une valeur ajoutée déterminante pour les activités de la société.

« *Les travaux 2022 de confiance.ai nous permettent d'ores et déjà de concrétiser la promesse d'une IA de confiance déployable en production. Sur plusieurs cas d'usage métier, nous avons pu en effet évaluer plusieurs paramètres de confiance tels que l'explicabilité et la robustesse au sein d'une chaine MLOps industrielle, prête à se conformer aux règlementations à venir telles que l'AI Act* », commente Yves Nicolas, Deputy Group CTO de Sopra Steria.

Enfin pour Renault, qui porte l'un des cas d'usage de référence du programme, c'est au travers la mise en application effective de premiers résultats que se porte l'intérêt du partenaire.

« L'enjeu de l'adoption et de l'intégration des solutions d'IA dans les systèmes industriels est un challenge d'autant plus important qu'il s'accompagne, pour les équipes Manufacturing de Renault Group, d'un changement de culture et de méthodes. Le programme Confiance.ai nous livre des outils clés en main, testés sur les cas d'usage industriels, proposés par nos équipes, qui nous permettent de consolider notre démarche globale de gestion des données industrielles. Aide à la qualité d'annotation, à la visualisation des données ou encore à la mesure de l'acceptabilité sociale de l'IA sur un poste industriel, les solutions proposées par les partenaires du programme Confiance.ai renforcent la robustesse de nos procédés et le temps d'exploitation de la donnée », explique Antoine Leblanc, expert AI @ industry 4.0 / DSII / PESI chez Renault Group.

Fin 2022 : Une deuxième version de l'environnement de confiance

Les travaux de l'année 2022 sont centrés sur les problématiques suivantes : montée en maturité de la robustesse, de l'explicabilité, du monitoring ou encore du cycle de vie de la donnée, caractérisation de la confiance, embarquabilité des composants d'IA, conformité aux référentiels. Par ailleurs, l'insertion de nouveaux processus d'ingénierie, en l'occurrence dédiés à la conception d'IA de confiance, dans les ateliers d'ingénierie mise en œuvre chez les industriels est un sujet d'étude central puisqu'elle garantit l'utilisabilité *in-fine* de l'environnement de confiance.

En 2022, l'environnement propose notamment quatre **plateformes** dédiées à des problématiques majeures de l'IA de confiance :

- Une plateforme consacrée à la gestion du cycle de vie de la donnée (acquisition, stockage, spécifications, sélection, augmentation)
- Un ensemble de librairies dédiées à la robustesse et au monitoring des systèmes à base d'IA. Elles permettent notamment d'assurer que le système et son composant d'IA évoluent bel et bien dans le contexte préalablement défini (Operational Domain Design).
- Une plateforme dédiée à l'explicabilité, dont l'objectif est de rendre en des termes compréhensibles par un humain les choix et décisions prises par une IA
- Et une plateforme destinée à l'embarquabilité des composants d'IA qui doit permettre d'une part, d'identifier sur la base des spécificités matérielles du système cible les contraintes de conception à respecter, et d'autre part, d'accompagner tout au long de la réalisation et ce, jusqu'au déploiement du composant dans le système.

Un important volet consiste en l'identification des critères de confiance pertinents (ex : fiabilité, robustesse, intégrité, pertinence, explicabilité, respect de la vie privée, etc.) pour un cas d'usage donné. Des modèles d'argumentation (**Assurance case**) sont proposés et permettent d'identifier les propriétés de confiance pertinentes, puis de définir comment valider le fait que ces propriétés sont atteintes. L'utilisation de ces différentes fonctionnalités sera facilitée au moyen d'un outil intitulé **« compagnon »** lequel instancie les méthodes construites dans le cadre du programme et

accompagne l'utilisateur métier dans le choix des critères de confiance, des niveaux de confiance à atteindre, des processus d'ingénierie à respecter et des composants à utiliser pour concevoir une IA de confiance.

A ce jour, **plus de 100 composants logiciels** (applications, librairies…) sont en en cours de conception dans le cadre du programme, à des niveaux de maturité différents. Progressivement évalués et intégrés, ils sont également mis à disposition des partenaires afin d'en permettre la manipulation dans leurs propres ateliers d'ingénierie.

En termes de rayonnement, **16 nouvelles publications** ont été acceptées sur le 1er semestre 2022 sur les différentes avancées du programme.

Prochaines étapes et perspectives

Confiance.ai adressera de nouveaux sujets tels que la **relation entre l'humain et l'IA,** suite à un appel à manifestation d'intérêt orienté sur les sciences humaines et sociales. Ont été retenues les propositions suivantes : Expérimentation de la confiance d'un utilisateur de système d'IA (IMS -ENSC Bordeaux), Cartographie de la situation morale : analyse de cas d'usage (LISN – CNRS), Le respect des valeurs de l'Union européenne by design par les systèmes d'IA (CDEP, Université Artois), et Les interfaces des systèmes algorithmiques : quelles informations communiquer pour générer la confiance ? (SCIA – Sorbonne Université).

La prochaine version de l'environnement sera centrée sur l'extension aux nouvelles approches issues de **l'IA symbolique et l'IA hybride**, en complément des approches orientées données en cours d'exploration. Une emphase particulière sera également portée sur les sujets « Méthodes et outils pour l'ODD » (Operational Design Domain) et les « approches end-to-end de l'explicabilité, la robustesse et du monitoring ».

De **nouveaux cas d'usage** seront proposés (une vingtaine au total) pour s'assurer de l'agnosticité des composants développés dans l'environnement de confiance.

Au-delà de sa participation au JTC21 (CEN-CENELEC) européen, le consortium Confiance.ai travaille également en partenariat avec **l'AFNOR** sur l'ingénierie des standards et plus particulièrement, l'émergence de « smart standards » face à l'augmentation en complexité du nombre de standards applicables.

Enfin, est étudiée la **pérennité de l'environnement** de confiance au-delà du programme, pour assurer sa continuité industrielle. Il sera proposé comme un ensemble modulaire à tous les industriels susceptibles d'en savoir besoin, au niveau européen et au-delà. Un schéma de valorisation est notamment en cours d'étude.

"*La démonstration de la confiance de bout en bout est un enjeu majeur pour assurer le déploiement de l'IA dans les systèmes critiques. C'est un des chantiers entrepris au sein de la deuxième version de l'environnement de confiance en appliquant un ensemble de travaux sur un même cas d'usage pour évaluer les complémentarités comme, par exemple la robustesse et le monitoring. Mais aussi nous avons abordé de nouveaux sujets comme le passage de la définition de l'ODD à la spécification des jeux de données* », commente Loïc Cantat, coordinateur technique du programme Confiance.ai.

« *La souveraineté des technologies numériques est au cœur des ambitions de France 2030. Nous devons à la fois protéger nos actifs et nos recherches, promouvoir nos valeurs mais également consacrer nos forces au développement d'une offre souveraine. Je constate qu'il existe déjà une vraie dynamique autour du programme Confiance.ai, qui réunit 50 partenaires déjà en mesure de proposer des solutions technologiques sur un marché estimé à 50 milliards d'euros. Ces travaux font de la France le leader de l'IA de confiance en Europe, c'est important de le souligner* », explique Bruno Bonnell, secrétaire général pour l'investissement, en charge de France 2030.

**Air Liquide, Airbus, Atos, CEA, Inria, Naval Group, Renault, Safran, IRT Saint Exupery, Sopra Steria, IRT SystemX, Thales, Valeo*

A propos de Confiance.ai

Porté par un collectif de 13 partenaires industriels et académiques français majeurs, Confiance.ai est le pilier technologique du Grand Défi « sécuriser, certifier et fiabiliser les systèmes fondés sur l'intelligence artificielle ». Lancé en janvier 2021 pour une durée de 4 ans, ce programme financé par France 2030 ambitionne de créer une plateforme souveraine, ouverte, interopérable et pérenne d'outils logiciels pour favoriser l'intégration de l'intelligence artificielle de confiance dans les produits et services critiques. Il fédère une quarantaine de partenaires industriels et académiques sur Saclay et Toulouse autour de 7 projets de R&D.

Confiance.ai est l'un des projets structurants sur lesquels s'appuie le second volet de la stratégie nationale pour l'IA, annoncé en octobre 2021. Ce programme est financé à hauteur de 30M€ par France 2030.

En outre, il contribue à la mise en œuvre par les industriels du futur règlement européen « AI Act » de la Commission Européenne.

Plus d'informations sur : https://www.confiance.ai/

A propos de France 2030

Le plan d'investissement France 2030 :

- ✓ **Traduit une double ambition** : transformer durablement des secteurs clefs de notre économie (santé, énergie, automobile, aéronautique ou encore espace) par l'innovation technologique, et positionner la France non pas seulement en acteur, mais bien en leader du monde de demain. De la recherche fondamentale, à l'émergence d'une idée jusqu'à la production d'un produit ou service nouveau, France 2030 soutient tout le cycle de vie de l'innovation jusqu'à son industrialisation.
- ✓ **Est inédit par son ampleur** : 54 Md€ seront investis pour que nos entreprises, nos universités, nos organismes de recherche, réussissent pleinement leurs transitions dans ces filières stratégiques. L'enjeu : leur permettre de répondre de manière compétitive aux défis écologiques et d'attractivité du monde qui vient, et faire émerger les futurs leaders de nos filières d'excellence. France 2030 est défini par deux objectifs transversaux consistant à consacrer 50 % de ses dépenses à la décarbonation de l'économie, et 50% à des acteurs émergents, porteurs d'innovation sans dépenses défavorables à l'environnement (au sens du principe *Do No Significant Harm*).
- ✓ **Sera mis en œuvre collectivement :** pensé et déployé en concertation avec les acteurs économiques, académiques, locaux et européens pour en déterminer les orientations stratégiques et les actions phares. Les porteurs de projets sont invités à déposer leur dossier via des procédures ouvertes, exigeantes et sélectives pour bénéficier de l'accompagnement de l'Etat.
- ✓ **Est piloté par le Secrétariat général pour l'investissement** pour le compte de la Première ministre et mis en œuvre par l'Agence de la transition écologique (ADEME), l'Agence nationale de la recherche (ANR), Bpifrance, et la Caisse des Dépôts et Consignations (CDC).

Plus d'informations sur : france2030.gouv.fr | @SGPI_avenir

Relations médias Confiance.ai
Marion Molina
Tél : 06 29 11 52 08
marionmolinapro@gmail.com

Secrétariat général pour l'investissement
Tél : 01 42 75 64 58
presse.sgpi@pm.gouv.fr