Data Science & AI

HUB Data Science & Artificial Intelligence

FEUILLE DE ROUTE 2022

L'année 2019 a été l'année de fondation des nouveaux Hubs et Enjeux du Pôle Systematic 4.0 dont le "Hub Data Science & AI" (DSAI).

Depuis, notre mission est de rassembler et d'animer un écosystème d'excellence, structuré autour des « sciences des données et de l'intelligence artificielle », avec comme ambition de devenir un acteur moteur et majeur de l'intelligence artificielle (IA) en France et en Europe. En 2020, le comité de pilotage du Hub DSAI a travaillé sur un premier plan d'actions comprenant des initiatives dans les sept dimensions de la vie d'un Hub du Pôle Systematic, étayée par une feuille de route. Ce document a pour objectif de présenter la mise à jour de cette feuille de route au regard des avancées scientifiques et techniques de ces deux dernières années avec une projection de leur TRL d'ici 2027 *[1]*.

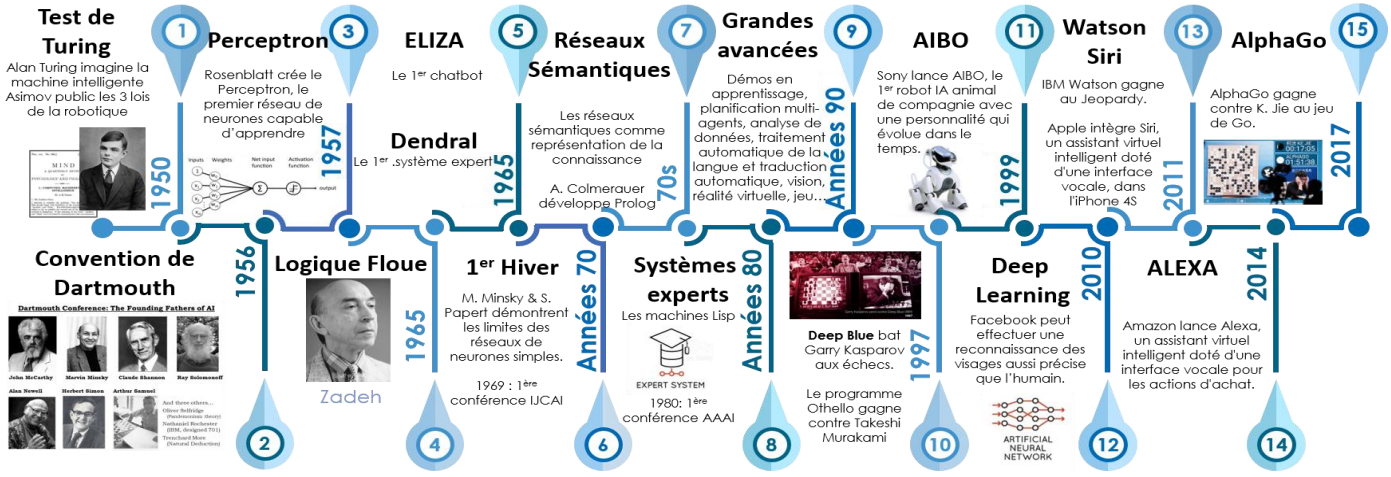
De plus, les prochaines années seront rythmées par de nombreux événements permettant de partager cette nouvelle feuille de route et de faciliter l'accélération de l'innovation et du développement des solutions à base d'intelligence artificielle et de sciences des données, issues de notre écosystème. Pour cela, nous proposerons : 2 Hub Days, des séminaires thématiques « Club DSAI » , une rencontre PME-Industrie, …. De plus, nous travaillerons en plus forte synergie avec les enjeux du pôle (Société, Industrie et Services et Territoires) ainsi qu'avec les autres hubs du pôle Systematic Paris-Region.

Vision Stratégique

L'IA aujourd'hui

Avec un marché mondial estimé à plus de 230 milliards d'euros en 2020 et dont la croissance atteindrait + 18 % par an pour atteindre plus de 450 milliards d'euros en 2024 *[2]* , l'Intelligence Artificielle (IA) est la tendance technologique de ce début de millénaire. Il est indispensable que les entreprises françaises soient bien positionnées sur ce marché pour pouvoir rester compétitives dans les années à venir. Le Pôle Systematic affiche clairement la thématique « Data Science & Artificial Intelligence » comme phare dans le cadre de la Phase 4 de la Politique Nationale des Pôles de Compétitivité.

Les systèmes d'IA basés sur les données se diffusent dans tous les pans de l'économie et de la société : grâce à eux, on peut mieux interpréter des radiographies et détecter des pathologies ; battre les meilleurs joueurs du monde au Go, au poker, à Starcraft ; déterminer le niveau de risque d'un prêt ou d'un investissement ; traduire la parole dans une quantité de langues ; superviser et optimiser des processus industriels ; produire des recommandations personnalisées de consommation ; et, prochainement, conduire des véhicules autonomes dans la circulation. Tout ceci est le résultat d'une longue histoire qui a débuté il y a plus de soixante ans.



[1]Les TRL (Technology Readiness Level) forment une échelle d'évaluation du degré de maturité atteint par une technologie. Cette échelle a été imaginée par la Nasa en vue de gérer le risque technologique de ses programmes. Initialement constituée de sept niveaux, elle en comporte neuf depuis 1995. Voir https://www.entreprises.gouv.fr/files/files/directions_services/politique-et-enjeux/innovation/tc2015/technologies-cles-2015-annexes.pdf

[2] Etude de marché réalisée par le SGPI avec le cabinet EY : https://www.ey.com/fr_fr/strategy/quel-avenir-pour-l-intelligence-artificielle-dans-l-industrie

Initiée lors de la deuxième guerre mondiale avec les modèles formels du neurone et les premiers travaux d'Alan Turing, formalisée en 1956 lors du congrès de Dartmouth, l'intelligence artificielle a connu des phases de développement enthousiaste et des phases de régression connus sous l'appellation d' « hivers de l'IA » : arrêt de la recherche sur les réseaux de neurones suite aux insuffisances techniques démontrées dans l'ouvrage de Minsky et Papert, Perceptrons (1969), puis arrêt des financements dans plusieurs pays suite au rapport de Sir Lighthill (1973) sur les promesses non tenues; remise en cause des investissements sur les systèmes experts dans les années 90 suite aux nombreuses limites rencontrées : problèmes de gestion de la cohérence, de passage à l'échelle, absence de connexion au monde réel et aux systèmes d'information conventionnels, problèmes de performance etc.

Ceci n'a pas empêché les chercheurs de développer de nouveaux algorithmes, de nouveaux outils, de nouvelles méthodes. C'est à la fin des années 1980, notamment, que les réseaux neuronaux convolutifs ont été conçus, accompagnés par des avancées importantes en matière algorithmique qui sont à la base de l'apprentissage profond et qui ont valu le Prix Turing 2018 à leurs concepteurs Yoshua Bengio, Geoffrey Hinton et Yann Le Cun.

Par ailleurs, le développement du web et l'accès rendu ainsi possible à des quantités de données de fonctionnement de la société et de l'économie a été catalyseur des importants progrès de la science des données et de l'IA depuis le début de ce millénaire ; c'est bien pour cela que les grands de l'IA et de la science des données sont aussi les grands opérateurs du web et de l'Internet : GAFAM aux USA et BATX en Chine, qui ont à leur service de formidables puissances de calcul pour soutenir ces développements.

Aujourd'hui, l'IA connexionniste connaît une très grande popularité, des investissements massifs, des salaires indécents pour ses meilleurs praticiens, en raison des performances atteintes par les systèmes d'apprentissage profond fonctionnant sur des supercalculateurs gigantesques et entraînés à partir de dizaines de millions d'exemples. Elle envahit tous les domaines de la société, mobilise des financements internationaux, effraie les populations, suscite des débats sur le futur du travail. L'IA symbolique, plus médiatiquement discrète, a cependant de nombreuses applications industrielles dans le domaine de la gestion des connaissances, de la logistique, de la planification, de l'automatisation des opérations ou de la résolution de problèmes complexes.

S'il est vrai que les systèmes d'IA sont performants sur des tâches individuelles comme la reconnaissance d'image, le jeu de Go, le diagnostic de maladie ou l'attribution de crédit financier, il est bon de modérer l'enthousiasme des aficionados et les propos alarmants tenus par certains prophètes technoscientifiques. D'une part, nul ne sait si les progrès technologiques de l'IA seront à la hauteur des espoirs formulés souvent sans justification. D'autre part, l'incapacité actuelle des IA entraînées par apprentissage à partir de données à expliquer en langage simple et adapté au contexte d'usage leurs décisions, et la très grande difficulté à prouver que ces décisions sont bonnes, pourrait engendrer de la méfiance, voire le rejet, de la part des utilisateurs.

Ainsi, qu'elle soit symbolique, connexionniste ou statistique, et/ou hybride, l'IA semble promise à un fort développement, il n'en demeure pas moins qu'un certain nombre de verrous ralentit son déploiement industriel, en particulier dans les domaines critiques comme la santé, la mobilité, l'industrie financière, les territoires intelligents, la sécurité et la défense, l'énergie… qui doivent par construction garantir des propriétés de sécurité et de sûreté mais aussi suivre des principes de confiance et de responsabilité. Aujourd'hui, de par le monde, un mouvement important existe pour développer des IA explicables ou, mieux, prouvables, voire certifiables et responsables. Une partie importante de la communauté IA mondiale a pris conscience de ce risque.

Éléments de contexte

Le ***Hub Data Science & Artificial Intelligence*** (DSAI) anime et fédère une communauté d'acteurs comprenant les académiques largement représentés dans le Pôle du fait de son ancrage territorial sur le Plateau de Saclay, mais aussi les industriels : startups, TPE, PME, ETI et Grands Groupes membres du pôle qui se positionnent massivement sur cette thématique. Ce Hub a aussi l'ambition d'accompagner cette communauté d'acteurs dans son développement dans une compétition d'ores et déjà mondiale et où la France et l'Ile de France ont un rôle à jouer.

En pratique, le Hub DSAI travaille sur un plan d'actions articulé autour de trois ambitions :

- Accompagner et accélérer le déploiement d'une intelligence artificielle de confiance,
- Faire progresser les algorithmes et méthodes,
- Valoriser les compétences de nos acteurs.

Pour décliner ce plan, le Hub DSAI collabore en étroite collaboration avec des partenaires tels que l'institut DataIA, les pôles de compétitivité NextMove, Finance Innovation, mais aussi l'IRT SystemX. Le pôle s'appuie aussi sur des collaborations avec d'autres acteurs nationaux (Teratec, Minalogic, le hub FranceIA…), européens et internationaux.

Le pôle Systematic en premier lieu, mais aussi les pôles dédiés uniquement aux usages et marchés tels que Cap Digital ou Finance Innovation, sont positionnés sur les domaines dans lesquels l'IA trouve de vrais enjeux : transport, sécurité, finance, industrie 4.0, santé, smart city, défense…

Systematic traite également des technologies très complémentaires à l'IA pour faciliter la création de plateformes de recherche, expérimentation et développement, notamment le Calcul Haute Performance.

Si l'éducation et la recherche sont fondamentales pour le développement de l'IA, le développement industriel, le succès commercial, avec à la clé la création de valeur et d'emplois, seront les signes de cette vitalité.

Le besoin de données / connaissances pertinentes et de bonne qualité mais aussi le partage de bonnes pratiques ou de retours d'expérience afin d'accélérer le déploiement de solutions opérationnelles à base d'IA, rend indispensable le lien avec les usages et les marchés, même dans les phases de recherche et développement.

L'animation des interactions entre académiques, industriels et utilisateurs finaux devient alors un facteur crucial pour le développement de cette technologie.

Axes technologiques prioritaires

Face aux enjeux portés par la science des données et l'IA, le Hub DSAI a choisi d'orienter son action sur six priorités technologiques au service des domaines d'applications :

- L'IA de confiance : la clé pour donner aux utilisateurs la maîtrise de ces technologies au moyen de preuves, de normes, d'explications, de garanties de sécurité ;

- L'évaluation des systèmes d'IA : une étape nécessaire pour garantir les performances de ces systèmes ;

- L'IA hybride : les liens entre modèles et données, apprentissage et modélisation, l'hybridation pour tirer le meilleur parti des données, des connaissances contenues dans les modèles des systèmes, mais aussi de modèles physiques (y compris venant de simulateurs) et les combiner ;

- Données structurées et non structurées, en stream (capteurs IoT) : parce que les grands volumes de données de demain proviendront des milliards de capteurs disposés dans les systèmes et dans l'environnement ;

- L'IA embarquée : parce que nombre de systèmes d'IA seront embarqués dans des dispositifs techniques de petite taille avec des contraintes de performance, de puissance, de consommation, de temps réel ;

- Les algorithmes d'apprentissage : parce que l'apprentissage ne peut se concevoir uniquement à partir de dizaines de millions d'exemples statiques, et que l'apprentissage actif, parcimonieux, non supervisé, sont des pistes pour concevoir les systèmes intelligents du futur.

Il faut cependant souligner que ces technologies impactent l'économie et la société dans toutes leurs dimensions ; les domaines d'application sont donc sans exclusive, la volonté du Hub DSAI est de couvrir l'ensemble des domaines applicatifs, en relation avec les enjeux identifiés par le pôle Systematic. Cette segmentation en 6 n'est pas aussi nette, car il existe de nombreuses dépendances entre chaque thématique. Par exemple, une évaluation maitrisée et expliquée des systèmes à base d'IA contribuera à augmenter la confiance. La constitution d'un jeu de données de bonne qualité (ex. Par réduction des biais) augmentera les performances d'un algorithme d'apprentissage…

Un accent particulier sera mis sur les programmes nationaux et internationaux ainsi que les plates-formes, notamment sur le programme Confiance.ai coordonné par l'IRT SystemX dans le cadre du Grand Défi National de de la certification de l'IA ou GaiaX. Systematic est par ailleurs membre de plusieurs projets Européens traitant de l'Intelligence Artificielle, tels que le projet REACH ou le projet DIH4AI.

REACH est un projet d'action pour l'innovation financé par l'UE lancé en septembre 2020. Il s'agit d'un incubateur Big Data de deuxième génération, qui s'appuie sur les efforts de l'EDI (European Data Incubator) pour accélérer l'innovation basée sur les données en Europe. REACH ira au-delà de l'EDI pour non seulement mettre en relation les entreprises de données avec les startups/PME, mais aussi pour engager les Digital Innovation Hubs (DIH) à développer des chaînes de valeur de données basées sur des données industrielles et privées propriétaires.

Enfin, il est clair que la diffusion de l'IA s'accompagne de considérations éthiques et réglementaires (AI Act), de standardisations, de respect de la vie privée, de questions de responsabilité et de transparence des algorithmes : les actions du pôle prendront naturellement en compte ces aspects.

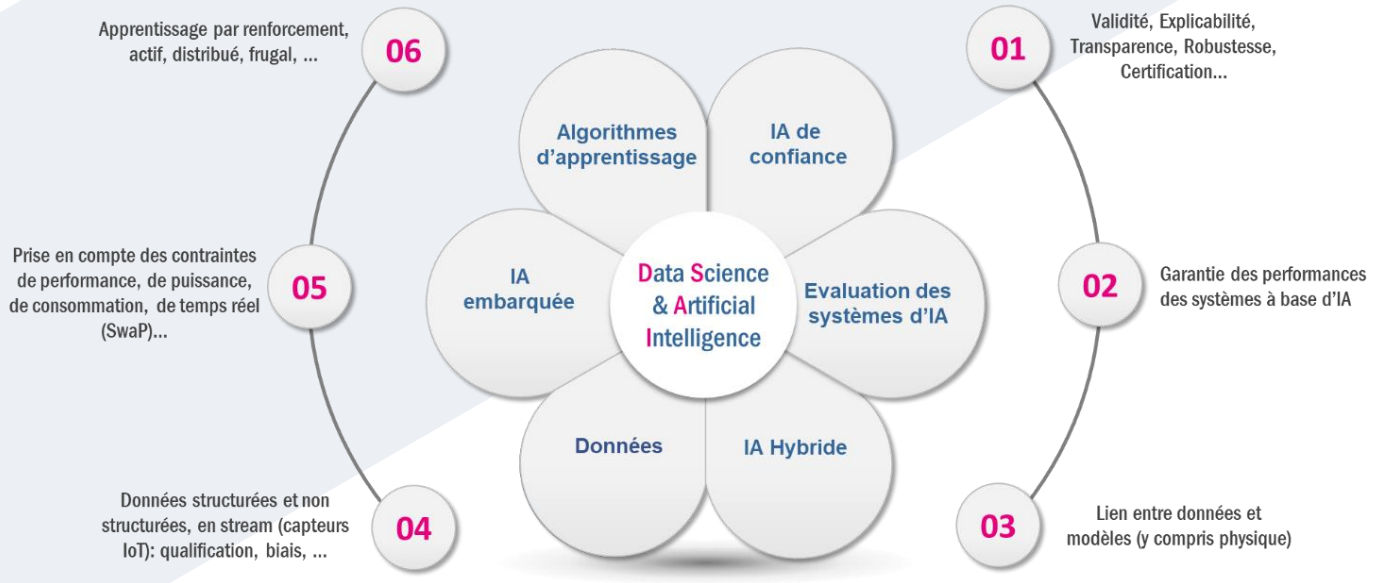
Positionnement national et international

Le Hub DSAI s'appuie sur la première communauté de recherche européenne en IA, selon le classement publié par Elsevier en décembre 2018, qui identifie l'Université Paris-Saclay comme le premier publiant européen en IA, devant les grands instituts britanniques et parisiens. Il s'appuie également sur une communauté économique de grands groupes, PME et startups qui représente une part conséquente des forces nationales en la matière. Il accompagne la Région Île de France notamment dans son Schéma Régional de Développement Economique, d'Innovation et d'Internationalisation (SRDEII) pour la région Ile de France sur la période 2022-2028 sur l'axe IA.

Saclay n'a pas été sélectionné pour y établir un des quatre instituts interdisciplinaires d'intelligence artificielle, mesure phare du programme national de recherche en IA ; on peut considérer ceci comme un accident de parcours, qui ne remet pas en cause la qualité des recherches et développements conduits sur le plateau, mais seulement la manière dont ils avaient été présentés à l'appel à manifestation d'intérêt. Autour de l'institut DATAIA et de l'IRT SystemX, la science des données et l'intelligence artificielle seront toujours des thématiques phares de l'action régionale, et la présence des grands industriels (Thales, Renault, Safran, Total, Atos, IBM …) garantit leur impact, aussi important que celui des futurs 3IA.

La compétition est évidemment féroce au niveau international et la France ambitionne d'être dans les cinq premiers acteurs mondiaux de l'IA et de la science des données, notamment grâce à son plan national d'IA résultat de la mission Villani. Le pôle Systematic, par son Hub DSAI, et par les relations qu'il entretiendra avec les autres Hubs et enjeux, sera un des acteurs de cette ambition.

Axes d'actions prioritaires



Les travaux du Comité DSAI de Systematic seront structurés autour des 6 axes prioritaires suivants :

- IA de confiance
- Evaluation des systèmes d'IA
- IA hybride, liens modèles et données
- Données
- IA embarquée
- Algorithmes d'apprentissage

IA de confiance

D'après le Journal Officiel du 9 décembre 2018, l'intelligence artificielle se définit comme étant le « champ interdisciplinaire théorique et pratique qui a pour objet la compréhension de mécanismes de la cognition et de la réflexion et leur imitation par un dispositif matériel et logiciel à des fins d'assistance ou de substitution à des activités humaines. » (Source : JO du 9 décembre 2018).

Ainsi, l'IA cherche à rendre un système informatique capable d'acquérir de l'information, de raisonner sur une situation complexe, de résoudre des problèmes combinatoires, de faire un diagnostic, de proposer une décision, un plan d'action, d'expliquer et de communiquer les conclusions obtenues, de comprendre un texte ou un dialogue en langage naturel, de résumer, d'apprendre, de découvrir. Mais, de plus en plus de systèmes à base d'IA peuvent avoir un impact sur nos vies. Ceux-ci embarquent des algorithmes pour assister les diagnostics médicaux, pour identifier les comportements anormaux dans une infrastructure critique, pour détecter des cyber-attaques... Les décisions publiques sont aussi concernées : lutte contre la fraude fiscale, attribution d'un logement social, affectation des élèves dans un établissement scolaire ou des étudiants à une formation... Cependant, qu'ils reposent sur des techniques d'apprentissage ou sur des approches plus symboliques, la conception de systèmes critiques à base d'IA n'est pas neutre et doit reposer sur une IA de confiance.

Qu'elle soit symbolique, connexionniste ou statistique, et/ou combinée à la science des données, l'IA semble promise à un fort développement, il n'en demeure pas moins qu'un certain nombre de verrous freine son déploiement en particulier dans les systèmes critiques, systèmes qui doivent par construction garantir des propriétés de sécurité et de sûreté mais aussi suivre des principes de confiance et de responsabilité. Mais, la conception de ces systèmes critiques n'est pas neutre. Elle doit reposer sur une IA de confiance et des ingénieries algorithmique et système adaptées et rigoureuses. Pour aller au-delà du PoC (Proof of Concept – Preuve de concept), il devient donc nécessaire de garantir des propriétés telles que l'explicabilité, la correction, la robustesse, … et la responsabilité, questions encore aujourd'hui ouvertes.

Verrous à lever

L'IA de confiance repose sur les cinq aspects suivants : l'explicabilité, la correction et la complétude, la contrôlabilité, la robustesse et la responsabilité.

• Explicabilité

Du point de vue de l'usager, le besoin est d'avoir une explication intelligible (explicabilité) plus que la traçabilité du raisonnement (interprétabilité). L'IA doit être capable d'expliquer, de façon intelligible, les raisons de ses choix/recommandations même si l'algorithme manipule, dans son fonctionnement, des notions ou concepts qui échappent à la compréhension humaine.

La qualité d'une telle explication peut être alors mesurée en fonction de son usage et de son destinataire, en s'inspirant des travaux de recherche en psychologie cognitive :

- Pour un développeur, l'explication doit lui permettre de comprendre le fonctionnement de l'application afin de le déboguer ou de l'améliorer ;
- Pour un utilisateur, l'explication doit lui permettre de comprendre le périmètre d'utilisation et les hypothèses sous-jacentes pour lui donner des clés de lecture sur les résultats obtenus ;
- Pour un expert, l'explication doit lui permettre de statuer un audit lors d'un incident.

De plus, la forme de l'explication doit pouvoir être évaluée et choisie pour minimiser le biais cognitif du destinataire.
Enfin, la transparence des algorithmes induit la capacité de l'algorithme à décrire précisément et de manière exhaustive ses mécanismes internes et à les interpréter de manière compréhensible pour l'ensemble de ses utilisateurs

• Correction et complétude

Un algorithme à base d'IA doit être tout d'abord valide c'est à dire qu'il doit faire ce qu'on attend de lui, tout ce qu'on attend et seulement ce qu'on attend. La preuve de correction doit apporter l'assurance que si l'algorithme termine, alors ses sorties sont correctes – c'est-à-dire qu'elles correspondent en fonction des données d'entrées à une solution au problème posé.
Cet aspect permet de vérifier la conformité entre les spécifications et le comportement du système à base d'IA, autrement dit l'écart entre ce qu'il est supposé faire et ce qu'il fait réellement. Certaines approches en IA symbolique comme la programmation par contraintes offrent, par construction, cette propriété de correction, mais il reste nécessaire de la démontrer dans les autres cas comme pour l'IA connexionniste.
On s'intéresse parfois aussi à la complétude qui garantit que l'algorithme donnera une solution pour chacune des entrées. Il est donc nécessaire de proposer des méthodes soit formelle, soit de validation empirique pour démontrer ces propriétés.

• Contrôlabilité

Dans de nombreuses applications et plus particulièrement dans le cadre des systèmes critiques, il est nécessaire de prouver que les algorithmes à base d'IA sont contrôlables, c'est-à-dire qu'ils sont bien-fondés ou cohérents (on emploie aussi l'anglicisme consistant). Ainsi, il faut démontrer qu'ils ne font que ce qu'on l'attend d'eux, interdisant ainsi des comportements émergeants non souhaités. Une approche possible est de contraindre les algorithmes d'IA à un domaine de viabilité des différents entrées/sorties et paramètres sous-jacents.

- **Robustesse**

La robustesse est une propriété plus difficile à garantir que la précision. En effet, un système non précis ne peut être robuste, et un système précis peut ne pas être robuste. C'est le cas d'un système à base d'apprentissage ayant appris par cœur les données d'apprentissage qui se trompera dans ses décisions futures basées sur de nouvelles données. Ainsi, il est intéressant de déterminer et mettre en place des solutions permettant aux systèmes d'IA fondé sur des notions d'apprentissage ou sur de l'hybridation d'IA, entre autres, d'être en mesure de garantir une certaine robustesse des décisions prises au cours du temps ou de remonter les déviances éventuelles des modèles sous-jacent.

- **Responsabilité**

La responsabilité des différents intervenants (acteurs qui conçoivent, développent, entraînent, maintiennent, contrôlent voire utilisent l'IA) doit être établie permettant une meilleure maitrise des biais intrinsèques acceptés ou involontaires.

Une monté en maturité pour atteindre TRL6 d'ici 2027.

La DARPA a lancé en 2016 son programme XAI (Explainable Artificial Intelligence) avec comme objectif principal de résoudre l'un des principaux défis de l'apprentissage automatique considéré aujourd'hui comme des techniques « boîte noire ». L'objectif du Hub DSAI est d'aller au-delà de l'explicabilité des approches neuronales ou d'apprentissage profond, proposant des solutions et méthodes pour pouvoir d'ici 5 ans déployer des applications à base d'IA de confiance, que celle-ci soit connexionniste, symbolique ou hybride.

En France, le conseil de l'innovation *[3]*, a lancé fin 2018 **le Grand Défi National "Sécuriser, fiabiliser et certifier des systèmes fondés sur l'IA"**, avec l'objectif de sortir de l'ère des PoC (preuves de concept) par une réponse appropriée à la question de la qualification (homologation voire certification) des systèmes critiques à base d'IA.

Il est donc nécessaire d'accélérer la recherche et l'innovation sur le sujet de l'IA de confiance pour atteindre un TRL6 d'ici 2027.
Les besoins sont de disposer :

- De méthodes et d'outils de gestion des données et des connaissances : conception, d'analyse, manipulation, collecte, acquisition, qualification, génération, filtrage des jeux de données d'apprentissage et base de connaissances pour la validation des systèmes cibles.
- De capacité à produire (concevoir, valider, implanter) un algorithme d'intelligence artificielle dit de confiance : correct, prévisible, stable, reproductible, explicable, fiable, robuste, capable de détecter les erreurs sur un domaine d'emploi défini et maîtrisé et donc in fine et si nécessaire certifiable.
- De capacité à définir et outiller l'intégralité du processus de développement, d'intégration et de qualification/certification sur l'ensemble du cycle de vie des systèmes intégrant de l'IA en interopérabilité avec les autres environnements de conception.
- De sortir d'une approche basée uniquement sur les preuves de concepts et passer à l'échelle industrielle en revisitant et repensant la chaîne d'ingénierie de l'algorithme, du logiciel et du système ainsi que la prise en compte du hardware pour le développement de composants à base d'IA.
- De nombreuses avancées autour de l'explicabilité pour les approches connexionnistes deviennent tangibles en particulier grâce au projet DEEL (DEpendable and Explainable Learning), programme de recherche franco-canadien, fruit d'une collaboration entre l'IRT Saint Exupéry (Institut de Recherche Technologique, Toulouse, France), IVADO (Institut de valorisation des données, Montréal, Québec) et le CRIAQ (Consortium de recherche et d'innovation en aérospatiale au Québec).

[3] Composé de 6 ministres, des administrations concernées (SGPI, DGE, DGRI), de deux opérateurs (ANR et Bpifrance) ainsi que de 6 personnalités reconnues, ce Conseil fixe les priorités stratégiques de la politique d'innovation française.

Les questions relatives aux problèmes de robustesse et de consistance commencent à faire l'objet de travaux liés aux preuves formelles. Ces dernières visent à apporter des garanties a priori sur la sûreté de fonctionnement d'un programme, contrairement aux méthodologies de validation par expérimentations directes qui visent à apporter des garanties a posteriori.

Plusieurs approches sont possibles : de la programmation par modèles (model checking), à la génération de code certifié, en passant par des langages intégrant directement les mécaniques de preuves, des solveurs logiques ou de l'interprétation abstraite de programmes. Ces approches montent progressivement en maturité pour atteindre le TRL6 d'ici à 3 ans.

Enfin, le programme Confiance.ai est un programme piloté par l'IRT SystemX, réunissant un collectif de 13 industriels et académiques français dont une grande partie est membre du pôle Systematic Paris-Région, va s'attacher à concevoir et à proposer une plateforme d'outils logiciels souveraine, ouverte, interopérable et pérenne permettant l'intégration de l'IA dans des produits et services critiques de manière sûre, fiable et sécurisée.

Connexion avec les autres Hubs et enjeux du pôle

Le Hub DSAI contribue à l'accélération du déploiement opérationnel, des méthodes outillées d'ingénierie des algorithmes d'IA (en lien avec le Hub "Digital Engineering") permettant d'aller au-delà de la preuve de concept (TRL3) et proposer des solutions/produits ou systèmes à base d'IA de confiance (TRL6) d'ici 2027.

Evaluation des systèmes d'IA

Contexte général

Pour des raisons tant économiques que réglementaires et sociétales, il est essentiel d'apporter des garanties quant à la sécurité et aux performances des systèmes d'IA. En effet, les systèmes doivent pouvoir fournir des preuves de leur conformité en vue de leur commercialisation. En outre, l'estimation des performances des systèmes peut constituer un avantage concurrentiel, via notamment l'identification d'axes d'amélioration ou de causes de sous-performance.

Dans un contexte européen qui voit l'émergence d'initiatives réglementaires et normatives tant pour l'IA que pour le Big Data et la robotique, il est nécessaire que le champ disciplinaire de l'évaluation de l'IA contribue à la pérennisation de la transition numérique de l'industrie. Dans ce cadre, deux thématiques prioritaires ont été identifiées. D'une part, dans le domaine de la mutualisation de la donnée (stratégies Open Data des collectivités, initiatives de partage de la donnée, etc.) : la donnée est centrale pour de nombreux systèmes d'IA et il est nécessaire d'estimer l'impact de la donnée « mutualisée » sur les performances de l'IA. D'autre part, il est nécessaire de poursuivre les efforts sur le développement de méthodes de référence pour l'évaluation de l'IA, qui permettront de maximiser l'adoption des systèmes d'IA dans la société.

Thématique : mutualisation de la donnée

- **Contexte**

La donnée est une composante essentielle des systèmes d'IA à base d'apprentissage. Elle peut également être impliquée dans la conception des autres systèmes d'IA, par exemple dans le cas où la constitution des règles expert s'appuie sur des analyses statistiques préliminaires du domaine. Dans un contexte encourageant la création d'un espace européen des données, il est nécessaire et profitable que la donnée soit mutualisable.

La donnée représente cependant un enjeu en termes de compétitivité économique face à la concurrence, de confidentialité, de propriété industrielle et de respect de la vie privée. Le développement de méthodes et bonnes pratiques pour le partage de données est essentiel, tels que la création de standards de formats maximisant l'interopérabilité des données (transverse au type d'implémentation algorithmique de l'IA, à la tâche, ou au domaine d'application), ou des procédures efficaces d'anonymisation et de pseudonymisation des données. Ces travaux liés à la mutualisation des données entraînent cependant une possible altération de la donnée, et il est nécessaire de garantir que la donnée « mutualisée » présente un niveau d'information adapté, et que la performance des systèmes d'IA est maintenue.

- **Quelles sont les contributions attendues en termes d'évaluation de l'IA ?**

- Développement de méthodes pour la qualification des données (représentativité, couverture, informativité, vérification des biais, etc.).
- Développement de méthodes pour l'estimation de performance des systèmes d'IA utilisant des données « mutualisées ».
- Détermination des facteurs d'influence de la performance des systèmes d'IA, notamment pour l'identification de phénomènes rares dans les bases de données.

Thématique : pérennité économique de l'IA

- **Contexte**

Un niveau TRL élevé, par exemple de niveau 8 ou 9, est souvent utilisé comme justification pour procéder au déploiement commercial d'un produit (système logiciel ou physique). Toutefois, l'échelle TRL ne concerne que la validation de la technologie, c'est-à-dire si celle-ci répond au cahier des charges initialement fixé, notamment en termes de performance des fonctionnalités et de domaine de couverture. La commercialisation et, éventuellement, l'industrialisation du produit nécessitent que le fabricant procède à différentes actions afin de réussir sa mise sur le marché. Ces actions sont principalement d'ordre économique et réglementaire, et peuvent être réalisées durant les phases de conception du produit (en parallèle de la montée en TRL) et suite à la conception (en post-TRL). L'évaluation de l'IA, dans ce contexte, ne repose donc pas simplement sur une estimation des performances fonctionnelles, mais bien sur la capacité du produit à être intégré d'un point de vue sociétal et économique.

D'un point de vue économique, il est nécessaire que le fabricant se soit assuré de la capacité de son produit à pénétrer le marché si le marché est déjà concurrentiel, ou de l'adéquation du produit avec un besoin consommateur s'il n'existe pas encore de marché pour ce type de produit. Il s'agira donc pour le fabricant, par exemple, de réaliser des études de marché, d'élaborer des stratégies marketing, d'analyser les besoins des consommateurs/utilisateurs finaux et de valider l'adéquation du produit avec ces besoins. Ces phases de développement d'une stratégie de commercialisation sont essentielles pour les entreprises souhaitant commercialiser leur produit, qui doivent s'assurer de sa viabilité économique dès les premières étapes de la conception. Cette capacité à appréhender le marché doit être accessible à toutes entreprises, afin notamment de maximiser la compétitivité des entreprises de petites tailles et des jeunes entreprises du secteur de l'IA, qui contribuent fortement à l'innovation et la transition numérique de tous les secteurs de l'industrie.

D'un point de vue réglementaire, le fabricant doit s'assurer que le produit est conforme aux dispositions applicables (en fonctions de la zone géographique, du la nature du produit, du marché visé, etc.). En cas de non-respect des exigences, les risques pour le fabricant sont bien sûr de nature légale et financière. L'évolution actuelle du cadre réglementaire va permettre au fabricant, dans les prochaines années, de disposer d'un ensemble clair et cohérent d'exigences pour les produits à base d'IA, qu'il s'agisse de solutions numériques ou de dispositifs physiques embarquant de l'IA (composants, capteurs, robots, etc.). Il est en premier lieu nécessaire que le fabricant soit en mesure d'identifier la réglementation applicable à son produit.

Un fabricant de robot industriel souhaitant développer et intégrer un composant de sécurité à base d'IA devra par exemple s'assurer de la conformité du produit à la réglementation européenne « Machines » (robotique et composants de sécurité) et à la réglementation IA. Le fabricant doit également être en mesure de procéder aux différents essais associés à l'évaluation de conformité réglementaire du produit (essais de robustesse, de durabilité, etc.), et de réaliser une analyse des risques couvrant de façon appropriée les capacités d'autonomie permises par l'IA. Il est donc nécessaire de poursuivre les travaux permettant aux entreprises de s'adapter au contexte réglementaire émergeant.

- **Quelles sont les contributions attendues en termes d'évaluation de l'IA ?**

- Initiatives de constitution de plateformes permettant la mise en réseau d'acteurs essentiels à l'accompagnement économique et réglementaire en lien avec les nombreuses initiatives, aujourd'hui très actives, dans la standardisation et la réglementation des systèmes à base d'IA.
- Définition de stratégies, méthodes, guides, permettant d'évaluer la performance du produit au regard des besoins des utilisateurs finaux (acceptabilité, adoptabilité et adoption du produit, utilisabilité, etc.) allant au-delà des activités classiques d'ingénieries et d'UX pour une prise en compte de l'adoption de solutions à base d'IA.
- Méthodes d'évaluation de conformité (performance, robustesse, résilience, explicabilité, transparence, etc.).
- Contributions au développement de normes et de référentiels de certification, par approches horizontale (IA) et sectorielle (métier, domaine).

Connexion possible avec les autres Hubs et enjeux du pôle

L'évaluation est une problématique transverse à tous les Hubs et tous les enjeux, car il est attendu de chaque technologie qu'elle fournisse des garanties, tant en termes de qualité que de conformité aux exigences légales, réglementaires, de standardisation et de sécurité.
Dans le cadre du Hub DSAI, l'approche de l'évaluation de l'IA devra être horizontale, et pourra également s'appuyer sur l'expertise sectorielle des autres Hubs.

IA hybride

L'apprentissage statistique a fait des progrès spectaculaires dans les deux précédentes décennies. Mais ceci ne doit pas masquer le fait que « l'intelligence » intégrée dans ces propositions reste somme toute assez sommaire. Les informations/connaissances injectables dans les phases d'apprentissage et de réglage des paramètres des algorithmes de machine learning (ML) se limitent le plus souvent à la fourniture de masse de données labellisées, dont seuls le volume et la qualité d'étiquetage permettent une amélioration des performances. Outre le coût prohibitif de la constitution de tels corpus quand il est possible de les obtenir, les applications pratiques ne disposent malheureusement que très rarement de base de données suffisamment grandes et exhaustives.

Pour continuer à progresser sur les challenges de classification, de diagnostic, de prédiction, de recommandation, … des systèmes industriels et des services aux usagers, il convient alors de se tourner vers des approches hybrides qui, en complément des données, prennent en compte d'autres types d'information.

Deux grandes catégories d'informations sont facilement intégrables : celles déjà encapsulées dans des modèles physiques (codes de calcul, outils de simulation, modèles d'état, …) et celles intégrées dans des modèles de connaissance et de raisonnement (ontologies, règles logiques, modèles sémantiques, …). Ces deux types d'information sont finalement les vecteurs principaux de capitalisation des ingénieries des grandes entreprises et leur prise en compte dans les nouvelles propositions d'IA permet d'associer les anciennes expertises avec les nouvelles en data science au sein des organisations.

A ce titre, le rapprochement des domaines calcul scientifique/simulation/modèle de raisonnement d'une part et ML d'autre part sont prometteurs et les challenges aux interfaces sont nombreux :

- Comment utiliser conjointement des données synthétiques et des données réelles dans un processus d'apprentissage ? Comment générer des données synthétiques complémentaires des données réelles disponibles ?
- Comment insérer un modèle physique dans un processus d'apprentissage (par la construction d'un méta-modèle) ? Comment valider le modèle hybride ?
- Comment ajouter de la connaissance experte (modèle sémantique) dans un processus d'apprentissage ?
- Comment contraindre le modèle d'apprentissage à respecter une structure issue d'un modèle physique ?
- Comment les techniques apprentissage peuvent aider à identifier des modèles physiques ?

Une monté en maturité pour atteindre TRL5 d'ici 2027.

La maturité de ces verrous est variable, mais la communauté scientifique s'en saisit fortement en ce moment. Sur un horizon de 5 ans, on peut espérer passer d'un TRL 3 à un TRL 5 sur ces sujets, possiblement 6 ou 7 pour certains d'entre eux comme celui de l'apprentissage utilisant à la fois des données de simulation et des données réelles (transfer learning).

Connexion avec les autres Hubs et enjeux du pôle

Ce thème de l'hybridation a des liens avec le thème 1 « IA de confiance (explicabilité, transparence, certification, vérification) » de la roadmap. Au-delà du Hub DSAI, il peut également avoir un certain écho dans le Hub "Digital Engineering".

A noter que ce thème sur l'hybridation est priorisé localement sur Saclay par plusieurs acteurs tels que l'Inria et l'IRT SystemX (programme IA2).

Données

Contexte technologique

"Les chiffres sont des êtres fragiles qui, à force d'être torturés, finissent par avouer tout ce qu'on veut leur faire dire", Alfred Sauvy.

Avec l'émergence du big data, des objets et de l'industrie connectés, de grandes quantités de données sont générées à chaque instant. L'importance de la qualité des données en machine learning et en IA est bien décrit par l'adage "garbage in, garbage out". Il signifie que tout modèle, malgré son degré de complexité, de performance ou de robustesse, voit sa faculté d'analyse borné par la qualité des données. De mauvaise données ne peuvent aboutir qu'à de mauvais modèles. De plus, pour les traiter en temps réel il est nécessaire d'élaborer des techniques d'analyse rapide sans avoir recours à l'ensemble des données passées. Dans ce cadre, le traitement de données non structurées (textes, vidéos) présente, dans certains cas, la difficulté supplémentaire de ne pas pouvoir reposer sur des catégories préétablies.

Verrous technologiques

Les principaux verrous technologiques sont :

- La caractérisation statistique du jeu de données, en particulier pour le cas de données non structurées ou non labélisées. Étape préalable indispensable au choix du modèle puisque tout dataset, aussi grand soit-il, reste un échantillon. Difficulté supplémentaire due au streaming. Niveau de TRL atteignable à 3 ans : 4-5 ; à 5 ans : 6 ; à 10 ans : 8.
- L'estimation et la prise en compte au sein du modèle de la qualité des données. Le processus doit identifier les anomalies, l'erreur potentielle de mesure (problèmes de métrologie), gérer les données manquantes. Niveau de TRL atteignable à 3 ans : 4-5 ; à 5 ans : 6; à 10 ans : 8.

- Dans le cas de données non structurées. Obtenir des indicateurs globaux pertinents synthétisant de grandes séries de données. Reconnaissance de profil (patterns) en temps réel. Automatisation de la reconnaissance de profil et adaptation des profils. Difficulté supplémentaire due au stream. Niveau de TRL atteignable à 3 ans : 2-3 ; à 5 ans : 4 ; à 10 ans : 6.
- Parallélisation des algorithmes. Étape nécessaire pour le passage à l'échelle sur la quantité de données. Lien avec le problème de l'automatisation du fenêtrage dans l'analyse des séries temporelles (comment partitionner le jeu de données ?). Difficulté supplémentaire due au stream. Niveau de TRL atteignable à 3 ans : 2-3; à 5 ans : 4; à 10 ans : 6.
- Élaboration d'algorithmes limités en mémoire. Informatique en périphérie de réseau (Edge computing). Niveau de TRL atteignable à 3 ans : 2-3 ; à 5 ans : 4 ; à 10 ans : 6.
- L'émergence de l'IA Frugale qui consiste à développer de nouvelles approches d'IA et de méthodes d'apprentissage automatique sur un faible jeu de données. (Fortement liée aux complications de procuration de données en Europe liée à la RGPD, à la rareté des données ou le coût de traitement de ses données).
- La confidentialité des données, le Federated learning pourrait constituer une solution à cette problématique lorsqu'on souhaite fusionner des données sensibles de deux sociétés différentes (RGPD ou exigences propres aux organismes). Plutôt que de centraliser les données pour y entrainer un algorithme central, l'apprentissage fédéré consiste à entraîner un algorithme sur la machine des utilisateurs d'une application et à partager ensuite les apprentissages ainsi réalisés.

Connexion avec les autres Hubs et enjeux du pôle

Ces enjeux sont reliés en particulier aux axes IA de confiance (obtenir des modèles transparents et explicables sur de grands jeux de données) et algorithmes d'apprentissage.

IA embarquée

Pour des usages liés entre autres à l'industrie du futur, le véhicule autonome, les objets connectés, la défense ou la cyber sécurité, le marché de l'IA embarquée apporte de réelles solutions sur le terrain et regorge d'opportunités. Ainsi, les algorithmes d'IA aussi bien à base d'apprentissage automatique que d'IA symbolique devenant opérationnels, un des grands enjeux est alors d'assurer la diffusion de ces logiciels au plus proche des structures embarquées à savoir les calculateurs SWaP (Size, Weight and Power) embarqués haute performance, l'IoT et le Edge, pour aller vers des systèmes IA pervasifs.

Il devient donc nécessaire de maitriser l'infrastructure embarquée supportant ces applications IA, infrastructures logicielles et matérielles de calcul, de communication et de mémorisation. Aussi, la capacité de pourvoir distribuer efficacement les applications à base d'IA sur ces infrastructures devient indispensable à maitriser. Enfin, le contrôle de la qualité de service de ces systèmes IA embarqués est également un verrou clé à lever, notamment en ce qui concerne la fiabilité, la sécurité ainsi que la prise en compte des contraintes de puissance consommée.

C'est pourquoi, une très bonne connaissance et voire une maitrise des composants matériels et logiciels permettant de concevoir des architectures de calcul embarqués pour l'IoT et le Edge sont clés et devront être à TRL5 au plus tard d'ici 5 ans. La capacité de développer des applications IA distribuées devient alors indispensable. Mais tout ceci ne pourra pas se faire sans des environnements de conception et de validation virtuelle des toutes ces nouvelles approches, environnements de développement capables de valider l'intégration des applications d'IA distribuées sur les modèles d'architectures embarquées spécifiques à chaque domaine d'application. Ces environnements devront être capables à terme de valider les performances, les propriétés fonctionnelles et non fonctionnelles de ces solutions en vue de leur certification.

De plus, la diffusion open source de ces environnements serait un atout majeur pour les industriels et le rayonnement du plateau de Saclay.

Algorithmes d'apprentissage

Description de l'axe

Les algorithmes et les applications d'apprentissage statistiques ont incroyablement progressé depuis une quinzaine d'année, tirés par le développement des techniques d'apprentissage profond et le renouveau des réseaux neuro-mimétiques. L'apprentissage statistique n'est, cependant, pas adapté à toutes les situations (*dans le désordre*) :

- Il requiert en général de gros volumes de données ; or, dans de nombreux cas d'application, les volumes nécessaires ne sont pas disponibles, ou pas avec les caractéristiques requises d'indépendance et d'identité de distribution. Ce peut être simplement parce que les données qui sont représentatives du problème à résoudre (au sens statistique du terme) sont intrinsèquement rares *[4]* ; parce que l'environnement d'application – et donc les caractéristiques du problème à résoudre ou la distribution des données – évolue trop vite par rapport à la quantité de données produites et que, en conséquence, la durée nécessaire pour accumuler suffisamment de données dépasse leur durée de vie utile *[5]* ; parce que le recueil des données est difficile ou coûteux et que, en conséquence, les données utilisables (de qualité suffisante, etc.) sont rares *[6]* (GAN, méthode générative, federated learning, data augmentation, …).

- Il prend difficilement en compte la connaissance existante si elle ne s'exprime pas sous une forme aisément assimilable par les techniques statistiques (données, fonctions de coût, distribution etc.), et, lorsque des connaissances a priori peut être intégrée dans l'apprentissage, il est rarement prouvable dans quelle mesure elles sont restituées par le modèle appris. Les modèles appris expriment, par construction, des corrélations statistiques, dont il est difficile de cerner, et plus encore de prouver, le domaine de validité, ce qui peut rendre problématique leur utilisation dans certains domaines *[7]* , et difficile leur portage ou l'extension de leur domaine d'utilisation (e.g. « transfer learning »).

- Il restitue la connaissance apprise sous forme de boîte noire, compliquant son acceptabilité par les utilisateurs, d'une manière générale, et son utilisation, plus particulièrement, dans les domaines où une explication peut être requise (dans le cas, par exemple, d'une décision ayant des conséquences pour l'utilisateur direct ou indirect), au point de la rendre quasiment impossible dans les domaines exigeant traçabilité et auditabilité des décisions (domaines réglementés).

- Environnement évolutif : La mise à jour des connaissances prend en compte le temps d'entrainement qui est le temps d'adaptation de notre IA à son nouvel environnement. Cette étape consiste à notre IA d'apprendre et de s'améliorer, il est donc indispensable d'examiner régulièrement les indicateurs de performance. Cette problématique est liée au temps de validité de notre modèle et des connaissances engrangées. On va donc chercher l'acceptabilité d'un modèle dans un temps raisonnable.

- Secteur sensible/critique : Le contrôle de l'IA dans les systèmes critiques, Il faut continuer d'y ajouter une intervention humaine pour contrôler la qualité du modèle. Jamais les machines ne pourront tout faire automatiquement, mais le processus peut être significativement accéléré. Le but principal est de réduire au maximum l'intervention humaine dans le processus d'apprentissage ainsi que de décision. Le danger serait que l'IA échoue dans ses tâches et repose essentiellement sur des efforts humains.

[4] On pense par exemple à la détection de fraude, aux données opérationnelles dans le domaine de la défense, aux maladies rares etc.

[5] Les données clients, marketing ou opérationnelles d'entreprises opérant sur des marchés en constante évolution.

[6] Recueil en milieu hostile, dépendant de technologie coûteuses ou peu fiables etc.

[7] typiquement parce que la redondance requise pour pallier à cette incertitude enlève l'essentiel de son intérêt à l'utilisation de l'apprentissage.

- Complexité algorithmique : elle réside derrière les algorithmes d'apprentissage automatique. La prise en compte de la limite pratique sur la capacité des ordinateurs à gérer la complexité des problèmes à traiter est primordiale. Le jeu de données, les variables explicatives/d'entrée, le degré de liberté du modèle et tous ces paramètres intrinsèques au modèle font contribuer à l'accroissement de la complexité.

- Contexte énergétique et écologique : Les ressources informatiques pour la recherche et des applications lourdes sont telles que l'IA a une empreinte carbone très importante. De meilleurs algorithmes et de méthodes d'apprentissage seront possibles grâce aux progrès en matière d'informatique quantique. Ne serait-il pas une solution de mettre des ouillères à la complexité algorithmique et d'y pallier grâce à l'évolution des ressources informatiques ? l'IA étant énergivore, on parle de pollution numérique et de l'impact environnemental qu'elle pourrait engendrer.

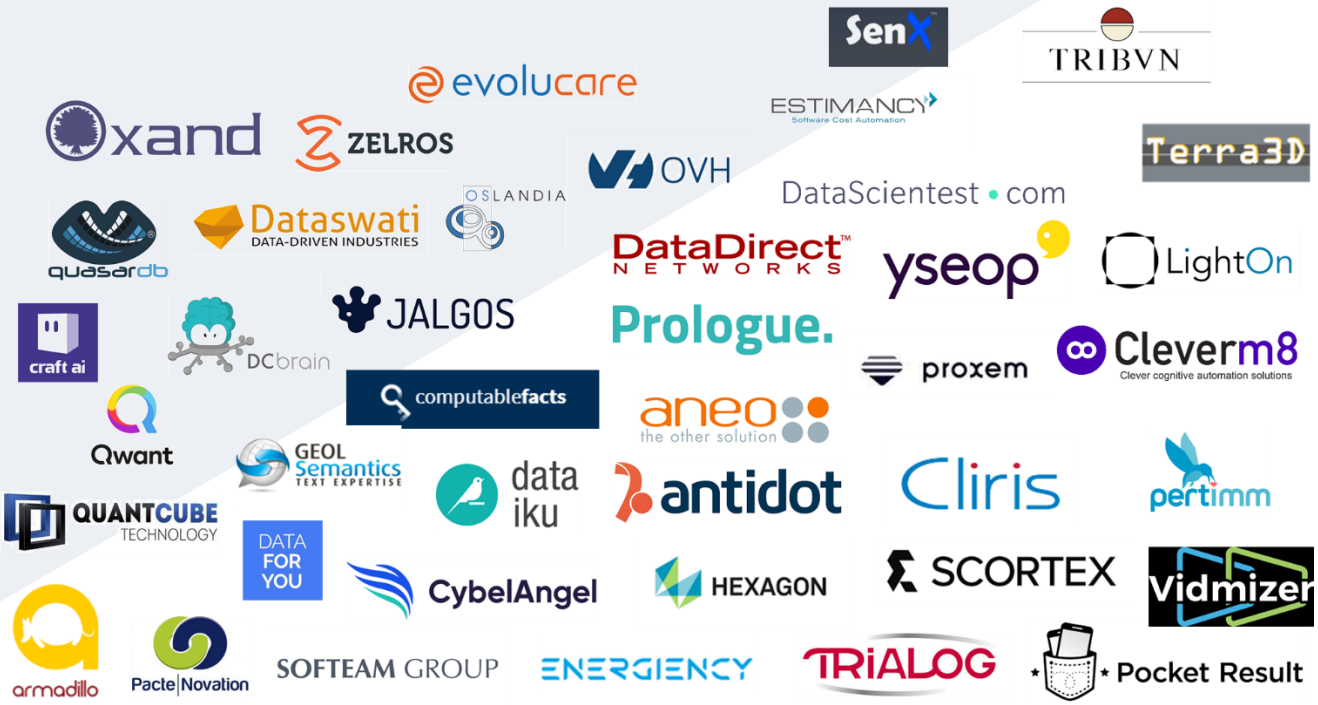
L'apprentissage symbolique n'a pas bénéficié de la dynamique créée par la nouvelle vague de l'apprentissage automatique et de l'IA en général, dans le sillage de l'apprentissage profond. On peut même dire que l'apprentissage symbolique a étonnamment peu évolué depuis la précédente vague (qui lui était plus favorable, il est vrai) : optimisations et améliorations à la marge, mais pas de progrès de fond. Pourtant l'apprentissage symbolique ne souffre pas des inconvénients listés plus haut pour l'apprentissage statistique, ou à un point beaucoup moindre : il représenterait même une solution possible dans de nombreux cas où ces difficultés disqualifient les approches statistiques. Mais il présente d'autre défis :

- Il est généralement peu robuste aux données bruitées.
- Il prend difficilement en compte la connaissance existante lorsqu'elle s'exprime sous une forme essentiellement statistique.
- Surtout, et c'est peut-être là la difficulté la plus fondamentale, les techniques d'apprentissage symbolique existantes ne traitent en général pas le problème du changement de représentation, lorsque le modèle conceptuel le plus pertinent pour représenter la connaissance à acquérir est différent du modèle des données d'entrée, ou ne le traite que dans le cas où la relation entre les deux modèles est connue.
- Tous les problèmes ne sont pas des problèmes de raisonnement, et il n'y a pas de raisonnement possible sans reconnaissance/identification des faits et situations

En fait, les faiblesses de l'apprentissage symbolique (manque de robustesse au bruit, apprentissage de représentation) correspondent assez directement aux forces de l'apprentissage statistique, et réciproquement : il est temps de combiner ces forces pour éliminer ces faiblesses.

Les acteurs du Hub Data Science & Artificial Intelligence

Les PME membres du Hub DSAI



Les Grands groupes membres du Hub

Les membres académiques du Hub

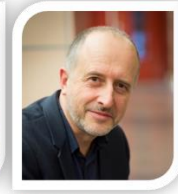
La Gouvernance du Hub

Présidente
Juliette Mattioli
Thales

Co-Président
Abdelhamid Mellouk
UPEC

VP Animation & Dév. Des Entreprises
Najah Naffah
Blockchain Secure

VP Développement des compétences
Caroline Chopinaud
Hub France IA



VP Recherche et innovation
Patrice Aknin
SystemX

VP Vision & Prospectives
Agnès Delaborde
LNE

VP Développement Marchés et Business
Matthieu Boussard
Craft ai

Responsable du Hub
Equipe Opérationelle
Johan D'Hose

Data Science & AI

Rédacteurs

Juliette Mattioli – Thales

Abdelhamid Mellouk – UPEC

Caroline Chopinaud – Hub France IA

Patrice Aknin – IRT SystemX

Agnes Delaborde – LNE

Matthieu Boussard – Craft ai

Christian De Sainte Marie – IBM

Johan D'Hose – Systematic Paris Region

Systematic
Paris Region Deep Tech Ecosystem