Vers une intelligence artificielle ubiquitaire explicable

Le développement d'IA ubiquitaires et explicables marque une rupture importante dans l'histoire de l'intelligence artificielle.

Il ne se passe plus une semaine sans que la presse ne rapporte les derniers progrès de l'intelligence artificielle ou ne mette en lumière des défis qu'elle s'apprête à relever. Certains articles publiés ont d'ailleurs parfois tendance à prêter plus de capacités à l'IA qu'elle n'en possède réellement, oubliant ainsi que l'état de l'art de l'IA est confronté à la complexité algorithmique des éléments.
Cela dit, la montée en puissance des techniques d'apprentissage automatique transforme en profondeur notre façon d'envisager le calcul, de construire des prévisions, de produire de l'aide à la décision, ou d'installer de l'autonomie dans les systèmes.

Nous ne nous situons qu'au tout début de l'histoire de l'IA et de cette ère de Turing qui trouve son origine dans les années 1950. Tout reste à faire ou presque pour déployer massivement des composants de *Machine Learning* (ML) qui rendront « intelligents » les objets de notre environnement et engendreront une intelligence artificielle ubiquitaire et explicable, accessible à tous, partout, tout le temps.

LES SIX DEFIS DE L'APPRENTISSAGE AUTOMATIQUE

La figure suivante représente les six principaux vecteurs de progrès attendus dans le domaine de l'apprentissage automatique qui permettront de déployer des IA ubiquitaires explicables.

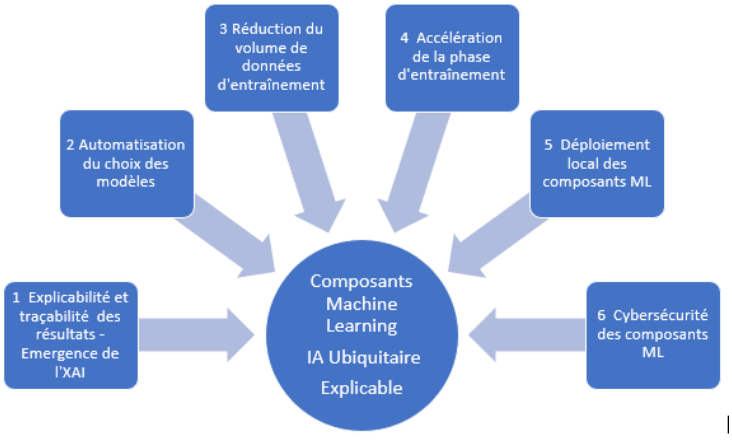

Figure 1 – Six vecteurs de progrès en Machine Learning

Figure 1 – Six vecteurs de progrès en Machine Learning

En réduisant les délais de mise en production des plateformes et les coûts de déploiement, les vecteurs de progrès 2, 3 et 4 vont rendre les composants ML de plus en plus performants et vont accélérer leur adoption dans l'ensemble des secteurs concernés, industriels, économiques, financiers et militaires. En abaissant la limite des volumes minimaux de données d'entrainement, l'apprentissage automatique adressera de nouveaux problèmes sur lesquels il demeure inopérant aujourd'hui faute de données suffisantes.

Le sixième vecteur de progrès s'applique à la cybersécurité des composants de Machine Learning. Les attaques par exemples contradictoires ou par orientation et influence de la phase d'entrainement d'une plateforme ML ont prouvé (en 2016) la vulnérabilité de ces architectures. Apporter de la sécurité « by design » à un composant ML contribuera à rendre résilient l'ensemble du système qui l'exploite.

Plus disruptifs encore, les vecteurs de progrès 1 et 5 donneront l'accès à de nouveaux domaines d'application et de développement pour les technologies du *Machine Learning.*
Le déploiement local de composants ML induit des problématiques d'optimisation des puissances de calcul et de stockage disponibles sur une architecture donnée, de compatibilité et d'intégration de ces composants à l'ensemble des systèmes, petits et grands. Comment en effet doter un simple objet connecté, commercialisé quelques euros, de capacités d'apprentissage sans augmenter sa puissance de calcul, ses capacités de stockage, de communication et donc son prix ? Comment installer des composants ML sur un réseau mobile, disposant de très peu de bande passante en utilisant des capacités de calcul distribuées parfois dégradées et sans s'appuyer sur une architecture centralisée ?

Ces questions, qui sont celles du déploiement de capacités d'apprentissage automatique en mode dégradé ou en mode *low cost*, conditionnent directement l'adoption des composantes ML à toute échelle et en tout lieu. Elles dessinent les contours d'une IA ubiquitaire, adaptative, totalement intégrée à l'environnement physique, d'accès simplifié et bon marché, disponible partout, tout le temps et pour tous.

L'INTELLIGENCE ARTIFICIELLE EXPLICABLE (XAI)

En tant que premier vecteur de progrès attendu, l'Intelligence Artificielle Explicable ou XAI (Explainable Artificial Intelligence) constitue la prochaine grande étape du développement des techniques d'apprentissage automatique. L'XAI va permettre de dissiper le brouillard algorithmique qui opacifie le mode de fonctionnement de nombreuses plateformes d'apprentissage automatique. Ouvrant la voie à la traçabilité et à l'explicabilité d'un résultat produit par un composant ML, l'XAI répond aux problématiques rencontrées dans des domaines aussi variés que le transport, la sécurité, le secteur militaire – défense, la médecine, la finance ou le juridique…

Aujourd'hui, l'utilisateur de composants ML qui obtient un résultat après calculs, souhaiterait pouvoir interroger le système pour comprendre d'où vient ce résultat. Ses principales questions seraient :

- Q1 Pourquoi fais-tu cela ?
- Q2 Pourquoi ne pas appliquer une autre méthode ?
- Q3 Quand ta méthode fonctionne-t-elle ?
- Q4 Quand échoue-t-elle ?
- Q5 Quand puis-je te faire confiance ?
- Q6 Comment corriger une erreur ?

Notons que l'explicabilité existe déjà par construction sur certaines architectures (arbres de décisions, random forest classifier). Le défi est de doter tous les dispositifs ML de capacités d'explicabilités *by design* parfois par ajout de nouvelles composantes ou fonctionnalités de traçage.

L'agence américaine de recherche appliquée à la Défense DARPA a lancé en 2017 son programme XAI avec l'objectif de généraliser le développement de plateformes ML explicables et d'abandonner toutes celles qui n'offrent pas cette transparence1. Le programme DARPA XAI établit un lien direct entre explicabilité, sécurité, confiance et résilience des composants ML et présente le concept XAI à l'aide des figures suivantes :

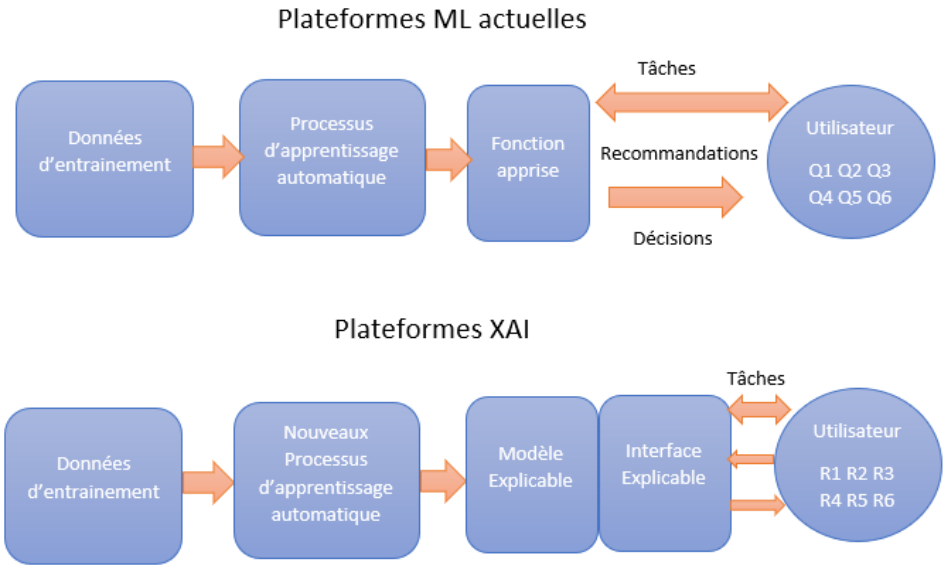

Figure 2 – Architectures ML actuelles et XAI

En utilisant une plateforme XAI, l'utilisateur obtient les réponses du système aux six questions d'explicabilité :

- R1 : Je comprends pourquoi tu fais cela
- R2 : Je comprends pourquoi tu ne fais pas cela
- R3 : Je sais quand tu réussis
- R4 : Je sais quand tu échoues
- R5 : Je sais quand te faire confiance
- R6 : Je sais pourquoi tu as fait fausse route

Favoriser le développement de plateformes d'IA explicable répond à trois exigences qui vont conditionner leur déploiement massif et leur adoption par l'utilisateur.

L'exigence de sécurité : l'utilisateur souhaite utiliser un outil qui ne soit pas détourné de ses fonctionnalités légitimes et qui préserve la sécurité de ses données personnelles.
L'exigence de confiance : l'utilisateur doit pouvoir accorder sa confiance à la plateforme, aux fonctions calculées et aux données de sortie qu'elle fournit.

L'exigence d'éthique : l'utilisateur attend d'une plateforme qu'elle assure un fonctionnement éthique et qu'elle s'interdise tout biais d'apprentissage

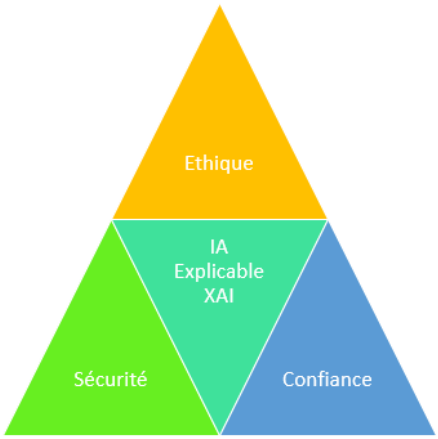

Figure 3 – Les apports de l'IA explicable

Figure 3 – Les apports de l'IA explicable

Le développement d'IA ubiquitaires et explicables marque une rupture importante dans l'histoire de l'intelligence artificielle. En répondant aux attentes de l'utilisateur, l'XAI participe à l'intégration de l'apprentissage automatique à l'espace physique tout en garantissant une forte lisibilité du mode de fonctionnement des plateformes. L'explicabilité engendre la confiance et permet d'éviter les biais analytiques. Elle favorise la sécurité et l'avènement d'IA éthiques, compatibles avec l'expertise humaine.

Thierry Berthier