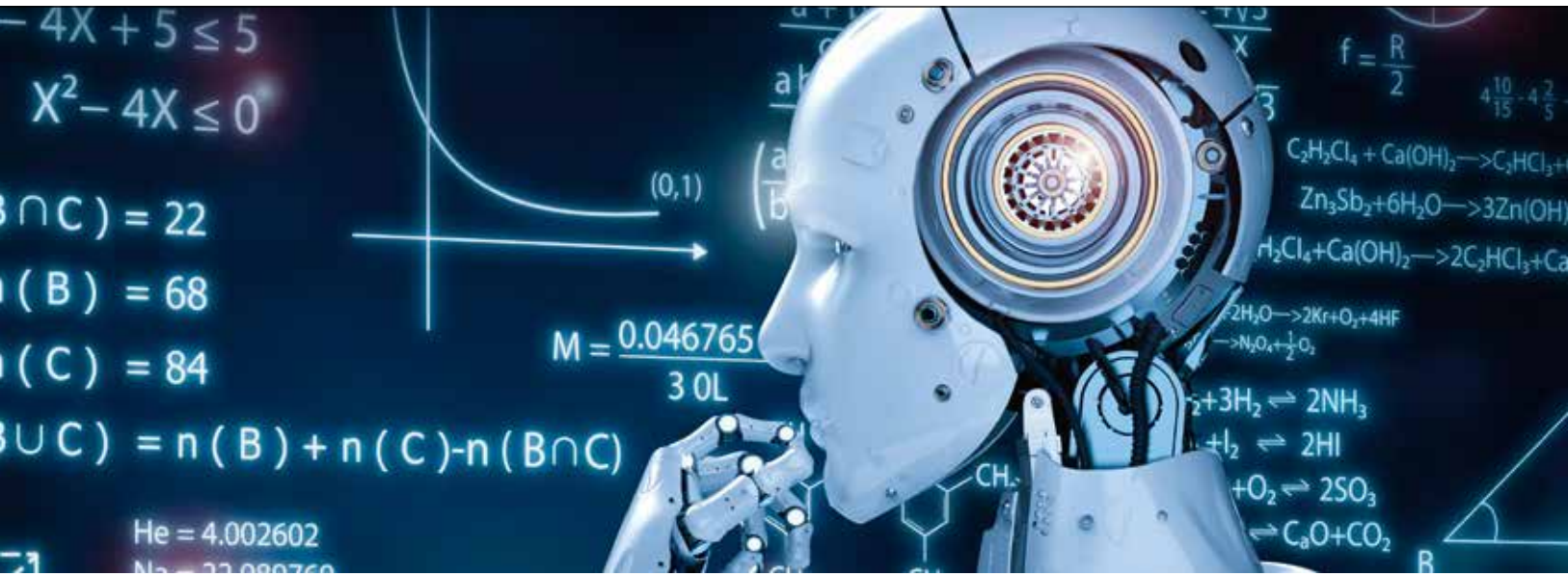
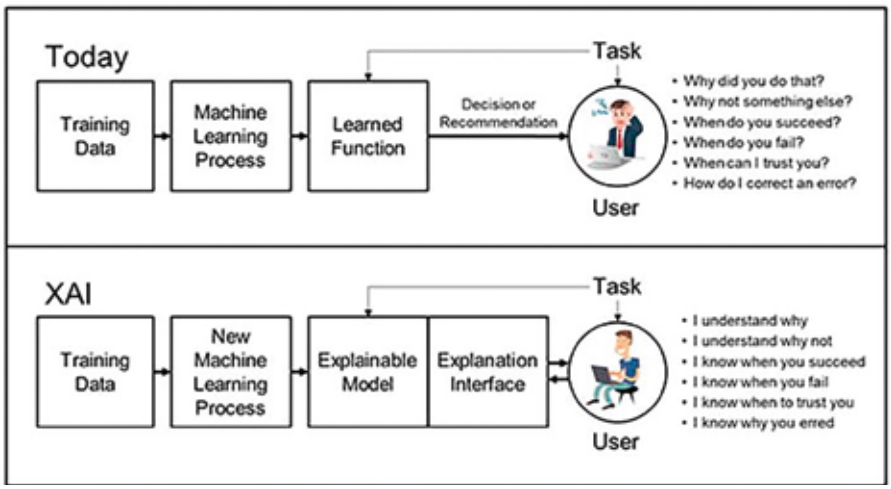


Une intelligence artificielle explicable est-elle la clé d'un développement ultérieur

L'intelligence artificielle est en plein essor. Même si nous ne le percevons pas toujours clairement, l'IA constitue pourtant un volet important de notre vie en société : dialogueurs, analyse d'images vidéo, recommandations sur Netflix, reconnaissance vocale ... Cependant, les décisions prises par l'IA soulèvent plusieurs questions importantes parce que les modalités de formation de ces décisions manquent parfois de clarté. L'IA explicable (en abrégé IAX) s'efforce de remédier à cette lasse incertitude.

PAR SAMMY SOETAERT

Sur LinkedIn, une fonction professionnelle intégrée transmet régulièrement des offres d'emploi à votre serviteur. Généralement, les postes vacants sélectionnés correspondent peu ou prou à mes compétences : concepteur-rédacteur, responsable de la communication, traducteur technique, etc. Dernièrement, j'ai même reçu une offre d'emploi portant sur un poste d'agent de surface dans ma boîte aux lettres électronique. Cette proposition a complètement raté sa cible et ne m'a valu que quelques ricanements à domicile. Je ne me suis pas borné à froncer les sourcils à la lecture de la proposition suivante, parce qu'à première vue, un poste de capitaine de port ne devrait pas me convenir. Je me demande comment diable LinkedIn en est subitement arrivé à la sélection de ces offres absurdes. LinkedIn a vraisemblablement estimé que les articles que j'ai rédigé ces dernières années portaient sur l'entretien et le nettoyage plutôt que sur la maintenance industrielle. Dans le même ordre d'idée, mes réactions fréquentes aux publications d'entreprises actives dans le secteur maritime pourraient suffire à expliquer l'offre d'emploi relative au poste de capitaine de port.



L'IAX permet aux utilisateurs de comprendre les algorithmes. (Photo Darpa)

Besoin de transparence dans le domaine de l'IA

Comme cette sélection bizarre n'avait aucun impact sur ma situation, je pouvais par conséquent me passer de toute explication concernant les mécanismes de décision sous-jacents. Mais supposons que votre banque fasse appel à l'intelligence artificielle pour déterminer

si vous pouvez prétendre à un prêt immobilier. Dans ce cas particulier, toute erreur d'appréciation se soldant par un rejet aurait des conséquences beaucoup plus sérieuses pour l'intéressé. Qu'en serait-il si l'intelligence artificielle embarquée sur des drones militaires considérait un véhicule comme un risque terroriste potentiel, alors qu'il ne s'agirait en réalité que d'une famille en train de faire ses emplettes ? Dans les situations comparables, il n'est pas toujours souhaitable que l'intelligence artificielle soit une boîte noire, cet endroit magique bourré de données et au sein duquel la mise à contribution de l'une des méthodes d'IA permet de générer un résultat. Une foule d'applications requièrent un processus de décision transparent. L'IAX est susceptible de répondre à cette nécessité. Mais outre les questions éthiques de cette nature, l'IAX est susceptible d'être mise à contribution dans le cadre d'autres applications, industrielles par exemple.

Climat de confiance

Professeur Bart Bogaerts est membre de l'Artificial Intelligence Lab mis sur pied par la VUB. Il nous en dit plus sur le comment et le pourquoi de l'IAX : « L'IAX ne s'arrête pas aux questions purement éthiques. Prenons par exemple le contrôle de qualité de pièces par le biais de la reconnaissance d'images. Un opérateur humain pourrait se demander pourquoi telle ou telle pièce a été exclue ? Dans certains cas, les raisons d'une telle exclusion ne sont pas toujours limpides, au risque de susciter une certaine défiance chez l'utilisateur. L'IA explicable pourrait dissiper cette défiance en communiquant les motifs sous-jacents de la décision prise. Dans le cas présent, il pourrait s'agir de la détection, sur une image vidéo, d'un certain nombre de pixels décolorés dont la présence serait révélatrice d'une anomalie. La capacité de l'utilisateur du système à appréhender le processus de décision qui anime l'IA pourrait contribuer à l'évaluation du système et à la confirmation de sa fiabilité. Je vais vous donner un second exemple justifiant l'adoption de l'IAX. Permettez-moi de rappeler un cas d'école lors duquel il a été demandé à l'IA de faire la distinction entre des photos de loup et de husky. Le logiciel est parvenu à atteindre un taux de réussite extrêmement élevé. Lorsque les observateurs se sont penchés sur les raisons d'une telle précision, ils n'ont pu en tirer qu'une conclusion décevante. L'identification précise de ces canidés n'était pas due à la reconnaissance de leurs caractéristiques individuelles, mais à la présence ou non de neige ou de verdure à l'arrière-plan. Sur la quasi-totalité des images de huskies, ceux-ci se dressaient sur un fond neigeux, alors que les loups avaient surtout été photographiés en milieu forestier. Par conséquent, l'IA s'était prononcée en se basant sur l'arrière-plan et non sur les caractéristiques de ces animaux. Un husky sur fond vert conduisait invariablement à un résultat erroné. Dans ce cas particulier, l'IA n'était pas parvenue à trouver la solution. En revanche, l'IAX présente l'avantage de rendre ces décisions intelligibles et dès lors d'en permettre l'évaluation. En cas de rejet d'un produit au cours d'un processus de production, l'IAX permet de se faire une idée plus précise des motifs de cette décision. Cet éclairage permet de rétablir la confiance dans le système et de mieux en appréhender le fonctionnement. En cas de doute, l'IAX est en mesure d'agir en quelque sorte sur le système. Ce levier permet de contrer ou non le système. »

S'agit-il d'un husky ou d'un loup ? L'analyse d'un algorithme fécond a révélé que les résultats exceptionnels obtenus étaient exclusivement imputables à l'arrière-plan et non aux caractéristiques des animaux. (Photo VUB)

Niveau d'explication souhaitable

La question se pose de savoir si, à terme, il ne sera pas impossible d'expliquer les décisions de l'IA. Les algorithmes ne cessent de gagner en complexité en raison du nombre croissant des variables introduites dans les systèmes que l'on observe, en particulier, dans l'industrie. Leur multiplication rendrait-elle plus difficile l'explication des algorithmes ? Bogaerts ne pense pas que la situation va devenir inextricable : « La question se pose de savoir à quelle profondeur il faut aller pour rendre un algorithme intelligible. Supposons qu'un véhicule autonome se déplace à une vitesse de 70 km/h et que vous vous demandiez pourquoi il en est ainsi à ce moment précis. Le renvoi à l'image d'un panneau de signalisation indiquant la limitation de vitesse correspondante suffira à vous convaincre. Il n'est pas indispensable de vous donner de surcroît une explication détaillée des raisons pour lesquelles l'algorithme pense que le panneau affiche le nombre 70, parce que cette donnée est immédiatement perceptible. Dans ce cas particulier, le niveau d'explication requis est évident, mais l'un des défis à relever réside dans la détermination du niveau souhaitable d'explication des algorithmes. Ce niveau dépend de la personne à laquelle on entend rendre un algorithme intelligible. » ■