



CIO.FOCUS

Intelligence artificielle

La comprendre et l'encadrer
pour en tirer des bénéfices

EN BREF

Si le concept d'intelligence artificielle date des prémices de l'informatique, son développement réel est bien plus récent. Il faut en effet de la puissance de calcul et du stockage en grandes quantités et à coût raisonnable pour pouvoir réellement y recourir. Aujourd'hui, l'intelligence artificielle s'immisce dans de nombreux processus d'entreprises. Face à cette fréquence croissante, de nombreuses questions apparaissent et une réglementation européenne s'annonce. Faisons ici un point.

Pour toute demande concernant CIO.focus :
contact-cio@it-news-info.com

Une publication de IT NEWS INFO :
48 Rue Cambon, 75001 Paris

Rédacteur en chef :
Bertrand Lemaire
blemaire@it-news-info.com
Tél. : 01 41 97 62 10

Principaux associés :
IT Facto et International Data
Group Inc.

Président et Directeur de publication :
Nicolas Beaumont

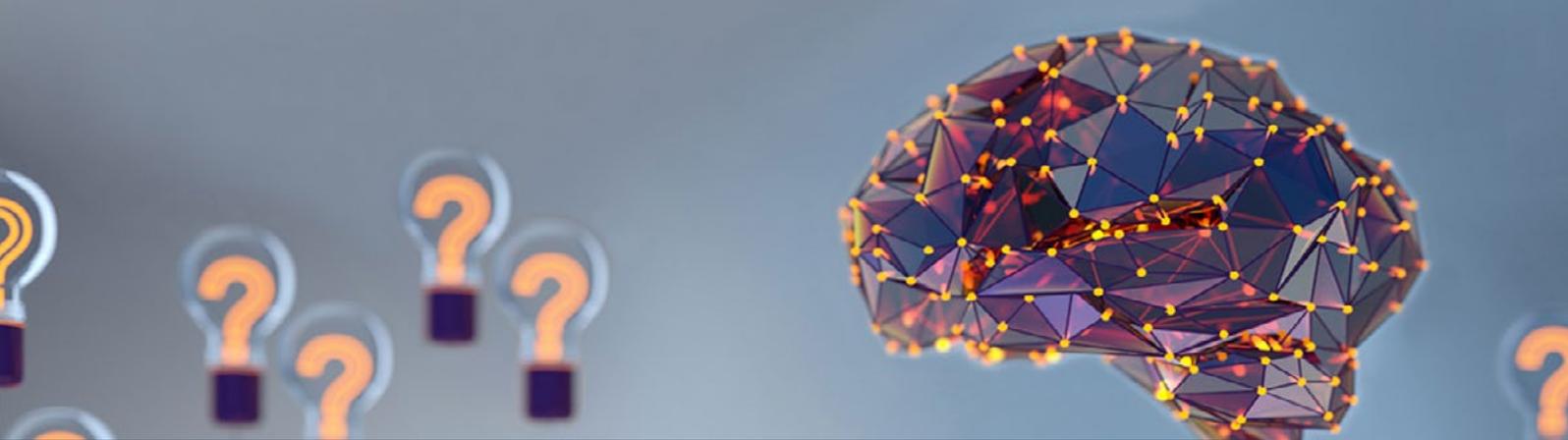
Directeur général : Nicolas Beaumont

CIO est édité par IT NEWS INFO,
SAS au capital de 3000000 €

Siret : 500034574 00029 RCS Nanterre

SOMMAIRE

/ TECHNOLOGIES	Pourquoi faut-il expliquer l'intelligence artificielle ?	3
/ TECHNOLOGIES	Comment expliquer l'intelligence artificielle ?	6
/ PROJETS	David Quantin (DGA OSII, Matmut) : « nous travaillons sur les usages avant de travailler sur les technologies »	11
/ TRIBUNE	Un cadre juridique européen pour l'intelligence artificielle : une première mondiale !	14
/ PROJETS	Comment La Banque Postale se prépare à rendre les algorithmes explicables	18
/ PROJETS	AG2R La Mondiale s'appuie sur l'IA pour accélérer le traitement des mails de ses clients	21



/ TECHNOLOGIES

Pourquoi faut-il expliquer l'intelligence artificielle ?

Pratiquement tous les domaines et secteurs d'activité déploient aujourd'hui des cas d'usage basés sur des technologies d'intelligence artificielle. Mais la fréquente opacité des algorithmes utilisés, en particulier quand il s'agit d'apprentissage machine, soulève en retour un besoin accru de transparence et d'explicabilité, surtout quand ces technologies ont un impact sur l'individu.

© Istock



Le développement des technologies de machine learning (ML) ou apprentissage machine a donné un nouvel élan à l'intelligence artificielle (IA). De la médecine aux véhicules autonomes, de la logistique à la cybersécurité, en passant par le monde militaire et policier, la finance, le marketing ou encore le recrutement : ces technologies d'IA sont aujourd'hui présentes dans tous les domaines. Selon les cas, elles servent à aider la prise de décisions d'un humain ou à déclencher des actions automatiques. Mais cet essor nécessite des garde-fous, en particulier quand les décisions prises ou guidées par l'IA ont des conséquences importantes sur les individus. Face à l'opacité de certains algorithmes, les individus ont en effet des droits, dont celui d'avoir une explication et de pouvoir contester les décisions les concernant. Mais cela suppose de comprendre comment un modèle d'IA fonctionne et comment il est arrivé à son résultat. Ce constat alimente un champ de recherche particulièrement dynamique à l'heure actuelle, celui de l'IA explicable, souvent abrégé en XAI (eXplainable AI).

Historiquement, les premiers modèles d'IA ont d'abord essayé de retranscrire des cheminements logiques sous forme de symboles. À la différence de cette IA dite symbolique, les modèles de machine learning, très populaires aujourd'hui, se construisent par l'apprentissage, à partir d'ensembles de données qui leur sont fournis en entrée. Il existe trois grandes catégories de machine learning. Dans l'apprentissage supervisé, le plus fréquent, des experts

humains indiquent à l'algorithme comment classer chaque élément de l'échantillon via des labels sur les données, jusqu'à ce que l'algorithme puisse inférer par lui-même le résultat. Avec le ML non supervisé, on laisse l'algorithme explorer les données par lui-même, pour qu'il découvre des corrélations ou des regroupements. Enfin, dans le ML avec renforcement, l'algorithme effectue des actions et apprend en fonction des réactions de l'environnement à celles-ci. Dans toutes ces approches, le modèle de décision n'est pas défini de façon formelle, mais induit à partir des données. Certains algorithmes de ML restent encore interprétables par un humain, dès lors qu'ils ne sont pas trop complexes, comme ceux basés sur la régression linéaire, la régression logistique ou les arbres de décision. D'autres en revanche s'avèrent plus difficiles à comprendre. Leurs résultats, qu'ils s'agisse de prédictions, de recommandations ou de classifications sont alors difficiles à expliquer. Certaines techniques de ML sont particulièrement opaques. Un guide récent publié par l'Information Commissioner's Office (ICO) et l'Alan Turing Institute cite en exemple les réseaux de neurones artificiels, les méthodes basées sur les ensembles, les machines à vecteurs de support et la forêt aléatoire, qualifiés de « boîtes noires ».

Protéger les droits fondamentaux des individus

Cette opacité soulève un certain nombre d'enjeux éthiques. Dès lors que les décisions basées sur des algorithmes d'IA concernent des individus, ceux-ci doivent en être informés. Ils sont en droit d'obtenir une certaine transparence, peuvent exiger de façon légitime une explication et contester la décision si jamais elle se révèle erronée ou injustifiée. Dans le cas où la décision est justifiée, l'expliquer permet aussi à l'individu d'agir pour avoir un meilleur résultat. Cela peut être par exemple un patient qui va modifier son alimentation pour diminuer un risque médical, un candidat qui va revoir son CV pour mieux mettre en valeur son expérience ou même une entreprise qui va retravailler son site Web pour améliorer son référencement. Plus les décisions touchent des domaines sensibles, plus les exigences de transparence et d'explicabilité sont généralement élevées. Plusieurs réglementations ont pour objectif de protéger les droits fondamentaux des individus face à l'usage de l'IA. C'est le cas notamment du RGPD, dont l'article 22 dit par exemple qu'aucun individu ne peut être sujet à des décisions entièrement automatisées ayant des conséquences légales, ou significatives, sauf si des conditions précises sont



© Istock

réunies : le droit d'obtenir une intervention humaine de la part du responsable du traitement, le droit d'exprimer son point de vue et celui de contester la décision. L'explicabilité risque de devenir encore plus essentielle avec la future directive européenne IA Act, qui distingue quatre niveaux de cas d'usage. Le projet de réglementation laisse en effet présager un net renforcement des obligations en vigueur, en particulier sur les cas d'usage à très haut niveau de risque.

Un autre enjeu clef est le risque de biais. La pertinence des algorithmes de ML dépend fortement de la qualité, de la quantité et de la diversité des données fournies, et les exemples ne manquent pas de situations où l'usage de l'IA a débouché sur des biais importants, pouvant entraîner des discriminations. Parmi les cas d'école figurent les algorithmes de reconnaissance faciale, qui il y a quelques années affichaient des écarts de performance notables selon le genre et la couleur de peau des individus, en raison d'un échantillon de photos où les hommes blancs étaient davantage représentés. L'explicabilité contribue aussi à déceler des biais éventuels, car elle se penche également sur la manière dont les modèles sont conçus et formés.

Une transparence associée à certains risques

En dehors de ces enjeux éthiques, l'explicabilité de l'IA apporte d'autres bénéfices. Pour les entreprises, offrir davantage de transparence sur la manière dont elles utilisent l'IA peut notamment augmenter la confiance des clients. Expliquer aide également à corriger et améliorer les modèles. Même dans le cas d'algorithmes ne touchant pas directement des individus, comme en météorologie, le fait de pouvoir expliquer les éléments qui interviennent dans les résultats peut permettre de comprendre pourquoi ces derniers sont parfois erronés.

Dans leur guide, l'ICO et l'Alan Turing Institute mettent toutefois en garde contre certains risques liés au fait de rendre les algorithmes d'IA explicables. L'un d'entre eux concerne le risque de voir l'IA détournée par des utilisateurs. S'ils en connaissent les mécanismes sous-jacents, ils sont alors en mesure de fournir les données pour arriver au résultat souhaité. Quand l'IA est utilisée par exemple pour la détection de

fraudes, il convient d'être davantage vigilant sur le niveau d'information fourni sur le fonctionnement des algorithmes. Un autre risque, plus classique, est d'ordre concurrentiel : quand l'algorithme apporte un avantage sur le marché à une entreprise, celle-ci peut hésiter à l'expliquer trop en détail. Enfin, le facteur du coût de l'explication entre dans la balance. Pour savoir à quel niveau l'entreprise doit investir, un rapport pluridisciplinaire coécrit par plusieurs chercheurs de Telecom Paris, Institut Polytechnique de Paris conseille d'évaluer le rapport risques-bénéfices de l'explication. Les auteurs indiquent en exemple qu'il y a peu de bénéfices à savoir précisément sur quoi Netflix base ses recommandations. En revanche, si une IA est impliquée dans un accident grave, l'explication offre alors un bénéfice élevé.

L'article suivant va aborder ce que signifie expliquer l'IA et les différentes manières d'y parvenir.

POUR ALLER PLUS LOIN

- [What is explainable AI?](#) Article de Violet Turri, assistante-développeuse dans le laboratoire d'intelligence artificielle du SEI (Software Engineering Institute), université Carnegie Mellon, 17 janvier 2022.

- [Explaining decisions made with AI](#), guide de l'Information Commissioner's Office (ICO) britannique et de The Alan Turing Institute, mai 2020.

- [Flexible and Context-Specific AI Explainability: A Multidisciplinary Approach](#), par Valérie Beaudouin, Isabelle Bloch, David Bounie, Stéphan Cléménçon, Florence d'Alché-Buc, James Eagan, Winston Maxwell, Pavlo Mozharovskiy et Jayneel Parekh, travaillant au LTCI, Telecom Paris, Institut Polytechnique de Paris et à i3, CNRS, Telecom Paris, Institut Polytechnique de Paris, 12 mars 2020



UN ARTICLE RÉDIGÉ PAR

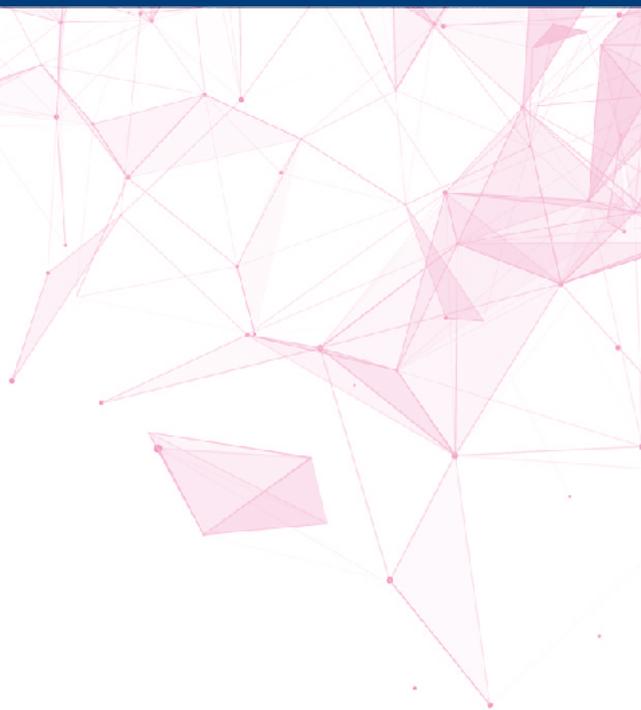
Aurélie Chandeze, Rédactrice-en-chef adjointe



/TECHNOLOGIES

Comment expliquer l'intelligence artificielle ?

Rendre les modèles d'intelligence artificielle compréhensibles devient de plus en plus important face aux nombreux enjeux soulevés par l'irruption de l'IA dans tous les domaines. Les entreprises ont tout intérêt à se pencher sur le sujet dès à présent. Pour bien démarrer, voici quelques axes pour engager la réflexion.



Après avoir fait un tour d'horizon des principaux enjeux de l'explicabilité de l'IA dans un précédent article, comment et par où aborder ce sujet pour le moins complexe dans son entreprise ou son administration ? Les différents rapports présentés en encadré proposent des réponses détaillées à cette question, dont peuvent s'inspirer les organisations. Cet article présente les grands axes de réflexion abordés dans ces travaux, en vue d'introduire et de sensibiliser au sujet.

Pour commencer, il faut rappeler que l'explicabilité va de pair avec la responsabilité. Les individus concernés par un traitement d'IA doivent savoir vers qui se tourner pour avoir une explication, ce qui signifie qu'un certain nombre d'acteurs dans l'organisation doivent endosser cette responsabilité. Selon le guide de l'ICO et l'Alan Turing Institute, la responsabilité globale de l'usage de l'IA est portée par l'équipe dirigeante, mais d'autres rôles sont concernés par le sujet. Parmi eux figurent les chefs de produit basés sur l'IA, les équipes qui développent l'IA, mais aussi les acteurs qui la mettent en oeuvre, en particulier les employés qui utilisent les outils. Enfin, l'équipe chargée de la conformité et le délégué à la protection des données (DPO) ont une responsabilité sur les aspects légaux.

Prendre en compte le contexte de l'explication

Ensuite, il faut comprendre qu'il n'y a pas une, mais de multiples explications possibles. Dans son guide, l'ICO énumère ainsi six



grands types d'explications. La première est l'explication rationnelle, qui présente le raisonnement aboutissant au résultat. Vient ensuite l'explication sur les responsabilités, indiquant qui a conçu quoi dans l'IA et qui contacter pour comprendre la décision. L'explication sur les données précise les données utilisées par le modèle d'IA et la manière dont elles ont été collectées, tandis que l'explication sur l'équité décrit les actions mises en place pour garantir celle-ci et éviter les biais. Il est également possible d'expliquer les mécanismes mis en place pour garantir la sécurité, la performance, la fiabilité et la robustesse ; de même que les mécanismes de surveillance de l'IA, pour évaluer ses conséquences sur les individus et la société. Il faut aussi faire la distinction entre l'explication globale, qui décrit le processus utilisé pour parvenir aux résultats et l'explication locale, qui détaille comment l'IA est arrivée à un résultat spécifique. Dans le cas d'un algorithme d'attribution de prêts bancaires, l'explication globale portera ainsi sur les données utilisées, les critères évalués et leur poids respectif, tandis que l'explication locale aura pour but de répondre à un client demandant pourquoi il n'a pas obtenu son prêt, en indiquant par exemple que dans son cas il dépasse le taux d'endettement fixé par la banque. Enfin, le rapport de Telecom Paris recommande de distinguer l'explication

sur l'algorithme en phase d'apprentissage, qui s'intéresse à la manière dont le modèle est entraîné, de celle sur l'algorithme entraîné et utilisé en situation opérationnelle, où l'on cherche plutôt à savoir les facteurs qui font qu'il arrive à tel ou tel résultat.

Pour produire une explication efficace, c'est-à-dire compréhensible et au bon niveau, mais aussi pour déterminer quelles explications fournir en priorité parmi celles décrites ci-dessus, le contexte est essentiel à prendre en compte. Les guides de l'ICO et de Telecom Paris détaillent tous deux les principaux facteurs contextuels, se rejoignant sur les grandes lignes. Ainsi, l'un des facteurs clefs est l'audience à laquelle l'explication est destinée, ainsi que son niveau d'expertise. Il peut s'agir de l'individu objet de la décision d'IA, comme le patient qui veut comprendre un diagnostic, l'utilisateur d'un service qui veut être rassuré sur le traitement de ses données ou le consommateur qui veut savoir pourquoi un prêt lui a été refusé. L'explication peut également s'adresser à des représentants de personnes morales, comme des entreprises ou administrations qui utilisent des solutions basées sur l'IA fournies par des tiers et ont besoin de transparence. Les instances d'audit, chargées de contrôler la conformité, peuvent bien

entendu être concernées. Enfin, en interne, les métiers et tous les rôles précédemment mentionnés peuvent avoir besoin d'explications.

De multiples facteurs déterminent l'explication à fournir

Il faut aussi tenir compte de l'impact de l'IA sur les individus, et en cas d'impact négatif, voir dans quelle mesure l'explication peut le diminuer. Les auteurs du rapport de Telecom Paris ajoutent une dimension liée aux réglementations applicables et aux droits fondamentaux, qui englobe le droit de contester les décisions, la prévention des discriminations, la protection de la vie privée et le fait de ne pas nuire à la société. Quand les cas d'usage de l'IA peuvent mettre en jeu la liberté, par exemple dans le domaine judiciaire ; l'accès à certains droits, comme l'éducation ou l'emploi ; voire entraîner des conséquences de vie ou de mort, comme dans la médecine, le secteur militaire ou encore les véhicules autonomes, alors les

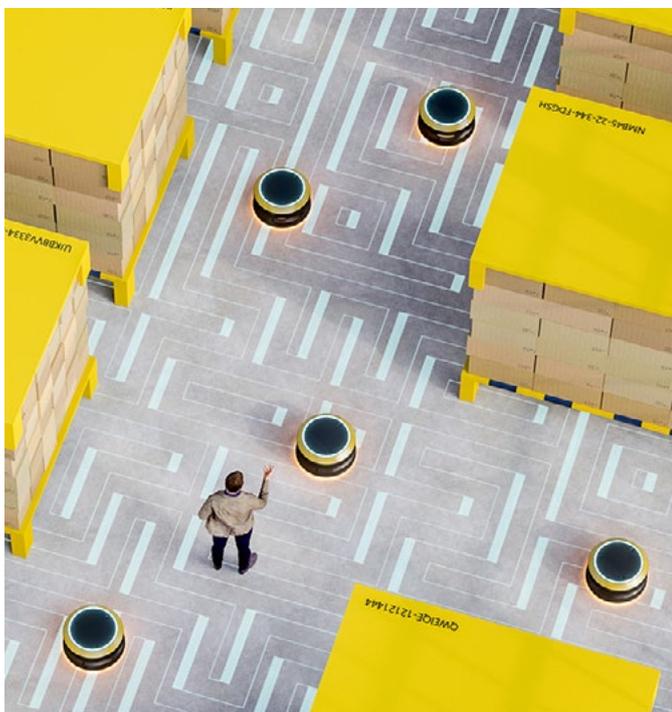
explications fournies doivent s'aligner sur les enjeux. Il s'agira par exemple de montrer ce qui est mis en oeuvre pour garantir l'équité dans un cas d'usage d'accès à l'enseignement supérieur, tandis que dans le cas d'un pilote automatique, les explications porteront en priorité sur la sécurité et la fiabilité.

Le contexte inclut également le domaine dans lequel travaille l'interlocuteur qui fournit l'explication : le niveau d'explication ne sera pas le même si celui-ci travaille dans l'e-commerce ou dans le monde médical. Les types de données utilisées par l'algorithme font également partie du contexte, car elles peuvent influencer le degré d'acceptation des résultats et les actions des individus, déterminant aussi le type d'explication à fournir. Typiquement, s'il s'agit de données sociales et démographiques, les préoccupations associées porteront davantage sur l'équité, alors que s'il s'agit de données médicales biologiques, l'enjeu est plutôt de rassurer sur leur fiabilité et la pertinence de la décision. Un autre facteur important pour ajuster le niveau d'explication est le degré d'urgence de la décision,



© Istock

autrement dit le temps dont disposent les individus pour recevoir l'explication et y réagir. Le rapport de Telecom Paris évoque enfin une dimension opérationnelle. Les organisations peuvent en effet avoir certains objectifs opérationnels qui requièrent d'expliquer les modèles d'IA, par exemple pour obtenir une certification de sécurité, ou tout simplement faire en sorte que les outils mis en place soient utilisés. Pour les dirigeants et managers, l'un des enjeux est ainsi de vendre l'IA auprès de leurs équipes afin d'obtenir leur adhésion. Pour les utilisateurs, expliquer facilite l'adoption en cas de réticence. Inversement, cela les aide aussi à prendre du recul, pour ne pas faire aveuglément confiance à l'IA et pouvoir détecter d'éventuels biais ou des résultats incohérents.



Un processus pour construire les explications

Une fois les différents paramètres contextuels identifiés, l'élaboration des explications peut démarrer. L'ICO propose six étapes pour y parvenir. Certaines demandent toutefois d'anticiper la question de l'explicabilité, pour l'inclure dès la conception du modèle d'IA. Tout d'abord, il s'agit de sélectionner les explications prioritaires en considérant le domaine, le cas d'usage et l'impact sur les individus. À ces aspects, le guide Telecom Paris ajoute une analyse des risques, en veillant à bien justifier les cas où la transparence doit être

limitée, ainsi qu'une analyse coûts - bénéfiques. Ensuite, il faut collecter et préparer les données en ayant en tête l'explicabilité. Le guide suggère pour cela d'utiliser le standard PROV-DM du W3C, un modèle de données décrivant leur provenance, à travers des relations entre des entités, des agents et des activités. La troisième étape consiste à construire le système d'IA de façon à s'assurer qu'on peut en extraire des informations pertinentes pour une diversité d'explications. Ensuite, il faut parvenir à traduire la logique du système d'une façon exploitable et compréhensible, en veillant à ne pas trop la déformer. Les auteurs du guide ICO mettent en garde en particulier contre le risque de transformer des corrélations, fréquentes avec les systèmes de ML, en causalités. La cinquième étape se penche sur la préparation et la formation des employés qui vont mettre en oeuvre l'IA. Enfin, la dernière étape s'intéresse à la manière de construire et de présenter l'explication.

De nombreuses méthodes existent ou sont en phase d'exploration pour expliquer les algorithmes d'intelligence artificielle. Selon les cas, elles peuvent fournir des éléments d'explication globaux ou une explication locale. Certaines aussi conviennent mieux à des sous-domaines de l'IA, comme les cartes de saillance (saliency maps) pour les algorithmes d'analyse et de reconnaissance d'image. Ces cartes affichent dans une image les pixels en fonction de leur degré d'importance sur le résultat. Enfin, il faut mentionner les méthodes d'explication post-hoc, utilisées quand il n'est pas possible d'accéder aux rouages internes du modèle d'IA. Ces approches sont parfois les seules possibles, notamment dans le cas des algorithmes en boîte noire. Elles parviennent à déduire certains attributs et la logique d'un modèle en le soumettant à diverses techniques, même si elles ont forcément des limites. Ces approches fonctionnent en général par perturbation, en fournissant différents ensembles de données en entrée et en regardant comment le résultat évolue. Elles aident à visualiser le fonctionnement du modèle, à comprendre le rôle et le poids des variables et comment elles interagissent et enfin à savoir ce qui devrait changer pour que la décision change (analyse contradictoire). Ces méthodes incluent par exemple LIME (local interpretable model-agnostic explanations), ICE (individual conditional expectation) ou les valeurs de Shapley, une méthode issue de la théorie des jeux.

Le guide écrit par l'expert en machine learning Christoph Molnar détaille les grandes méthodes utilisées à l'heure actuelle.

En dehors de ces méthodes, un autre champ d'études envisage d'intégrer l'explicabilité directement dans les modèles d'IA, pour avoir une IA explicable « by design ». Des approches d'IA hybrides sont notamment évaluées pour rendre les algorithmes plus transparents, sans pour autant renoncer aux techniques avancées de ML là où celles-ci sont performantes. On réduit ainsi la partie opaque des modèles. Il s'agit par exemple d'associer des règles et du ML dans un modèle d'analyse d'image, pour délimiter la zone d'une image où le ML doit s'appliquer. Comme autre exemple, le guide de Telecom Paris cite une règle disant qu'un robot ne doit pas s'approcher « trop près » d'un humain, combinée à du ML pour établir selon contexte ce qu'est « trop près ».

Expliquer n'est pas justifier

Pour conclure, il est intéressant de mentionner les propos de Daniel Le Métayer et Clément Hénin, respectivement directeur de recherche et docteur en informatique, qui ont travaillé ensemble au sein de l'équipe Inria Privatics. Interviewés par le laboratoire d'innovation numérique de la CNIL (LINC), leur regard permet de prendre un peu de recul sur ce domaine en pleine effervescence de l'IA explicable. Ils rappellent notamment qu'expliquer n'est pas justifier, en distinguant quatre notions : la transparence - le fait de rendre visible le fonctionnement de l'IA ; l'explication - le fait de rendre compréhensible ; la justification - pour rendre la décision acceptable et enfin la contestation, pour convaincre qu'une décision est mauvaise. Se contenter de fournir une explication ne suffit pas, il faut s'assurer d'être en mesure de la justifier, et quand la décision concerne un individu, permettre à ce dernier de la contester. Ensuite, Clément Hénin pointe un travers fréquent à l'heure actuelle, le fait que beaucoup d'outils d'explicabilité soient fournis par ceux-là mêmes qui conçoivent les solutions d'IA. « Ce domaine souffre encore beaucoup de ce que le chercheur Tim Miller et ses collègues ont appelé le syndrome des détenus qui veulent diriger l'asile [...], en l'occurrence des experts d'IA expliquant leurs propres productions », observe

ainsi Clément Hénin. Il en résulte des explications figées et unilatérales, qui ne répondent pas forcément aux besoins des différents utilisateurs. Avec en suspens, un autre enjeu, « mesurer la qualité d'une explication ou d'une justification ».

POUR ALLER PLUS LOIN

Voici également différents documents en accès libre sur lesquels cet article s'est appuyé.

- [Explaining decisions made with AI](#), guide de l'Information Commissioner's Office (ICO) britannique et de The Alan Turing Institute, mai 2020.

- [Flexible and Context-Specific AI Explainability: A Multidisciplinary Approach](#), par Valérie Beaudouin, Isabelle Bloch, David Bounie, Stéphan Cléménçon, Florence d'Alché-Buc, James Eagan, Winston Maxwell, Pavlo Mozharovskiy et Jayneel Parekh, travaillant au LTCI, Telecom Paris, Institut Polytechnique de Paris et à i3, CNRS, Telecom Paris, Institut Polytechnique de Paris, 12 mars 2020

- [Explainable Ai \(XAI\): A Systematic Meta-Survey Of Current Challenges And Future Opportunities](#), méta-analyse par Waddah Saeed et Christian Omlin, Centre de recherche en IA de l'université d'Agder Grimstad, Norvège.

- Clément Henin et Daniel Le Métayer : « [Fournir des explications du fonctionnement des algorithmes compréhensibles par des profanes](#) » | LINC (cnil.fr). Entretien avec Clément Henin et Daniel Le Métayer, respectivement doctorant et directeur de recherche au sein de l'équipe Inria Privatics, réalisé Félicien Vallet, 7 janvier 2021

- Thèse de Clément Hénin : [Expliquer et justifier les systèmes de décisions algorithmiques](#), février 2022

- Pour avoir un aperçu détaillé des différents outils et techniques existants pour rendre le machine learning interprétable : [Interpretable Machine Learning, A Guide for Making Black Box Models Explainables](#), par Christoph Molnar, 21 février 2022



UN ARTICLE RÉDIGÉ PAR

Aurélie Chandeze, Rédactrice-en-chef adjointe

/ PROJETS

David Quantin (DGA OSII, Matmut) : « nous travaillons sur les usages avant de travailler sur les technologies »

La Matmut s'est lancée dans des premiers usages de l'intelligence artificielle en s'appuyant sur la start-up Zelros. Cette initiative correspond à une stratégie générale du groupe mutualiste de recours à des start-ups pour éviter d'internaliser des technologies avant d'en avoir démontré la valeur d'usage.

© DR



© Matmut

David Quantin, directeur général adjoint organisation, systèmes d'information et innovation de la Matmut, a obtenu chez Zelros ce qu'il était venu y chercher.

Née en 1961, la Matmut (Mutuelle d'assurance des travailleurs mutualistes) est un groupe d'assurances mutuelles dont le siège se situe à Rouen. Ses 6 400 collaborateurs servent 4 millions de sociétaires avec 7,6 millions de contrats d'assurances (IARD, santé, automobile, professionnelle, vie...), d'épargne, de retraite, etc. au travers de la maison-mère et de plusieurs filiales. « La Matmut revendique un positionnement d'assurance 360° » martèle David Quantin, directeur général adjoint organisation, systèmes d'information et innovation de la Matmut. Le groupe produit l'essentiel de ses assurances mais peut être amené à compléter son offre avec des partenaires, justement pour avoir une couverture totale du marché, par exemple en distribuant de l'assurance retraite de BNP Paribas Cardif. De même, s'il a une logique de distribution essentiellement directe multicanale, il s'appuie aussi sur Cdiscount comme distributeur, cette enseigne constituant un « carrefour digital » pour aller à la rencontre des clients. Pour innover ou améliorer son service, la Matmut n'hésite donc pas à s'appuyer sur des tiers. C'est notamment le cas avec des start-ups afin d'intégrer des technologies non-encore maîtrisées en interne. Pour l'intelligence artificielle, la Matmut a ainsi choisi de s'appuyer sur Zelros.

Si la Matmut a toujours pratiqué la vente à distance, le groupe mutualiste a depuis plusieurs années largement digitalisé sa relation client. David Quantin relève : « nous ne sommes plus dans une logique de digitalisation de la relation client, question réglée depuis



longtemps. Depuis 2015, les clients peuvent effectuer la totalité de la gestion de leurs contrats en ligne et la souscription est aujourd'hui totalement digitale dans 25 % des cas pour la santé, 5 à 6 % pour l'IARD. » Le recours à l'intelligence artificielle a été identifié comme un vecteur de performance notamment dans cette phase de distribution.

Des cas d'usage identifiés

« La chaîne de valeur de l'assurance comprend trois étapes principales : la conception des produits, la distribution et enfin la gestion du contrat et des éventuels sinistres » indique David Quantin. Il ajoute : « l'intelligence artificielle peut contribuer à plusieurs endroits dans cette chaîne de valeur et nous avons notamment pensé à l'automatisation de tâches, à la découverte/apprentissage sur la connaissance client et enfin sur l'augmentation du conseiller ou du gestionnaire. Dans un premier temps, nous avons voulu augmenter les conseillers de la Matmut dans la phase de distribution. » Le but est donc que le conseiller se voit assisté dans sa relation par une intelligence artificielle pour lui suggérer des actions. Selon les données disponibles sur un client, le gestionnaire va, par exemple, pouvoir lui suggérer des offres complémentaires (aide à la vente en fonction de l'appétence à telle ou telle offre

selon les données disponibles) ou mieux apprécier le risque avant souscription. David Quantin note : « le projet est orienté sur l'expérience collaborateur ».

Mais, pour atteindre l'objectif assigné à l'intelligence artificielle, encore faut-il disposer des technologies nécessaires. « Autant nous aimons maîtriser et faire nous-mêmes, autant nous cherchons des accélérateurs lorsque nous avons besoin de technologies que nous ne maîtrisons pas encore » explique David Quantin. Il justifie : « plutôt que de passer du temps à acquérir des compétences sans savoir quel sera le retour réel, nous travaillons sur les usages avant de travailler sur les technologies. Pour cela, nous nous appuyons sur des technologies packagées, notamment issues de start-ups, avant, en cas de succès, d'internaliser le savoir-faire. » La Matmut a procédé de la sorte pour son chantier data et a appliqué la même méthode pour l'IA.

Une stratégie favorisant les start-ups

Depuis 2015, la Matmut a en effet une stratégie délibérée de travailler avec des start-ups voire d'y investir via un véhicule financier dédié, Matmut Innovation. Bien entendu, le groupe mutualiste a comme volonté prioritaire de repérer les start-ups opérant dans le secteur de l'assurance. La Matmut est de ce fait membre de Le Swave, la partie fintech, assurtech et regtech de l'incubateur Paris&Co. C'est dans ce cadre que la Matmut a d'abord rencontré Zelros. Puis cette start-up a également participé à une

« *Zelros est une start-up prometteuse et dédiée au marché de l'assurance. Leur catalogue technologique est orienté autour de notre métier.* »

présentation dans un hub de découverte organisé par la Fédération Française d'Assurance qui a donné une visibilité à Zelros auprès des assureurs français. David Quantin juge : « Zelros est une start-up prometteuse et dédiée au marché de l'assurance. Leur catalogue technologique est orienté autour de notre métier. »

De janvier à septembre 2021, la Matmut a expérimenté

la solution de Zelros. « Ce n'était pas un démonstrateur mais bien une expérimentation avec de vrais clients, un sous-ensemble de nos sociétaires, et certains de nos conseillers » insiste David Quantin. Pour cette phase d'expérimentation, les données issues du CRM Salesforce et les données du SI de gestion déversées dans la plate-forme data sous Saagie en on premise n'étaient pas déversées en direct dans Zelros mais poussées ponctuellement. Delphine Mennereuil, responsable de pôle projet à la Matmut, explique : « nous avons débranché en septembre 2021 cette première implémentation, une fois la valeur confirmée et les axes d'amélioration identifiés. »

Après l'expérimentation, la généralisation

Actuellement, avec une mise en production programmée pour courant 2022, la Matmut travaille sur l'implémentation définitive. La première différence sera un lien direct temps réel entre la plate-forme data et Zelros. « Et comme nous voulons un produit orienté expérience collaborateur, nous travaillons avec des UX designers pour nous assurer d'une bonne intégration aux processus de travail et une interface appropriée aux usages » ajoute Delphine Mennereuil. David Quantin confirme : « nous ne voulions pas passer du temps à construire toute une pile IA avant d'avoir travaillé sur les usages mais, maintenant, la démonstration de valeur a été faite ».

Concrètement, le collaborateur opère à partir de son écran Salesforce. Zelros prodigue des recommandations dans une pop-up (argumentation...). Le conseiller peut ensuite réaliser un feed-back sur les propositions de Zelros dans le but d'en améliorer la performance. La solution, de ce fait, peut entamer un auto-apprentissage. « Les contenus poussés dans la pop-up sont créés et gérés en interne car nous voulons en conserver une totale maîtrise et garantir un discours unique face aux sociétaires » précise Delphine Mennereuil. Enfin, la Matmut calcule des indicateurs clés de performance de la solution.

Des perspectives encore à définir

Installé localement, Zelros a donc clairement apporté la valeur attendue par la Matmut. David Quantin se

souvent : « nous n'avons pas eu à former d'experts en IA mais nous sommes montés en compétence sur le sujet. Avec Zelros, nous avons obtenu ce que nous étions venus chercher chez eux. Nous réfléchissons actuellement, au-delà de l'actuelle généralisation, à d'autres cas d'usage. Nous travaillons également en parallèle sur les cas d'usage liés à la reconnaissance et la lecture automatisée de documents. »

Si le budget global n'est pas précisé, le coût de l'expérimentation initiale a été qualifié de « faible coût habituel pour ce genre de test » par David Quantin. Quant à l'actuelle généralisation, la Matmut est sorti du cadre de licence habituel de Zelros pour adopter un accord de type open-bar sur ses technologies en échange d'un montant forfaitaire et d'une approche partenariale. David Quantin se réjouit : « cela nous permet d'avoir une relation sereine, gagnant-gagnant voire même plutôt donnant-donnant, et de réaliser un travail commun sur nos cas d'usages. »

POUR ALLER PLUS LOIN

- 23 Avril 2018 : [Grâce à Michelin, la Matmut teste la conduite améliorée](#)

- 30 Mai 2018 : [David Quantin devient DGA en charge des SI à la Matmut](#)



UN ARTICLE RÉDIGÉ PAR

Bertrand Lemaire, Rédacteur en chef de CIO



/TRIBUNES

Un cadre juridique européen pour l'intelligence artificielle : une première mondiale !

Les avocates Christiane Féral-Schuhl et Justine Sinibaldi détaillent ici le projet de réglementation européenne, basée sur une approche par les risques, en matière d'IA.



© Istock



© Lika Banshoya

Christiane Féral-Schuhl est avocate associée du Cabinet Féral-Schuhl / Sainte-Marie.

À l'issue de deux années de travaux très denses, la Commission européenne a présenté fin avril 2021 sa proposition de Règlement en matière d'intelligence artificielle [Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union du 21 avril 2021, COM(2021) 206 final, 2021/0106(COD)]. Ce texte, actuellement en discussion au Parlement européen et au Conseil de l'Union européenne, constituerait, s'il est adopté, une première mondiale, répondant à l'objectif de faire de l'Union européenne « un acteur mondial de premier plan dans le développement d'une intelligence artificielle sûre, fiable et éthique » [Résolution du Parlement européen du 20 octobre 2020 contenant des recommandations à la Commission concernant un cadre pour les aspects éthiques de l'intelligence artificielle, de la robotique et des technologies connexes, 2020/2012(INL)].

Le sujet n'est pas nouveau puisque, dès 2017, le Parlement européen avait envisagé la création d'une personnalité juridique spécifique aux robots autonomes. Si cette voie a été abandonnée, la Commission européenne a réaffirmé sa volonté d'encadrer les risques associés à l'utilisation de l'IA, tout en favorisant son développement au sein du marché européen. La présidente Von der Leyen a ainsi annoncé dans ses orientations politiques pour la Commission européenne « une proposition législative en vue d'une approche européenne



coordonnée relative aux implications humaines et éthiques de l'intelligence artificielle » [Orientations politiques pour la prochaine Commission européenne 2019-2024, Mme Ursula von der Leyen].

De nombreux travaux ont ainsi été menés au cours des deux dernières années, tant par le Parlement européen [Par ex. résolution du Parlement européen du 20 octobre 2020 concernant un cadre pour les aspects éthiques de l'intelligence artificielle, de la robotique et des technologies connexes, 2020/2012(INL)] que par la Commission européenne. Celle-ci a mené une vaste consultation des principales parties intéressées, y compris des universitaires, entreprises, partenaires sociaux, organisations non gouvernementales, experts de haut niveau sur l'IA ainsi qu'une analyse d'impact des différentes options stratégiques possibles en termes d'intervention réglementaire (labellisation non obligatoire, approche sectorielle ...).

C'est en fin de compte une approche proportionnée fondée sur les risques qui a été retenue par la Commission. Elle consiste en un traitement gradué en fonction du niveau de risques présenté par le système d'IA concerné afin de répondre au double objectif affiché par la Commission européenne [Livre blanc de la Commission intitulé « Intelligence artificielle - Une approche européenne axée sur l'excellence et la

confiance », COM(2020) 65 final, 2020] : promouvoir le recours à l'IA et tenir compte des risques associés à certaines utilisations de cette nouvelle technologie.

Tout d'abord, à l'instar du RGPD, la proposition de règlement prévoit des règles d'application extraterritoriale : (i) aux fournisseurs établis dans l'Union européenne, mais aussi à ceux établis dans un pays tiers qui mettent sur le marché des systèmes d'IA dans l'Union européenne (ii) aux utilisateurs de systèmes d'IA situés dans l'UE, et (iii) aux fournisseurs et utilisateurs de systèmes d'IA situés dans un pays tiers, dès lors que les résultats générés par le système sont utilisés dans l'UE.

Ensuite, trois niveaux de risque des systèmes d'IA sont définis, obéissant à des régimes différents.

« *Les pratiques d'IA présentant des risques inacceptables car contraires aux valeurs de l'Union sont interdites.*

Les pratiques d'IA présentant des risques inacceptables car contraires aux valeurs de l'Union sont interdites. Il s'agit notamment (i) des systèmes d'IA altérant le comportement des personnes par le

recours à des techniques subliminales ou exploitant les vulnérabilités liées à l'âge ou au handicap d'un groupe de personne (ii) du recours par les pouvoirs publics à des systèmes d'IA évaluant ou classant la fiabilité des personnes en fonction de leur comportement social, ou encore (iii) de l'utilisation de systèmes d'identification biométrique à distance « en temps réel » dans des espaces accessibles au public à des fins répressives. Cette liste pourrait toutefois être étendue suite à l'avis conjoint du Comité européen à la protection des données (EDPB) et du Contrôleur européen de la protection des données (EDPS) du 18 juin 2021 [Avis conjoint 05/2021 de l'EDPB et du CEPD sur la proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées sur l'intelligence artificielle (législation sur l'intelligence artificielle), 18 juin 2021] qui estiment notamment que toutes les techniques d'IA de notation sociale ou de reconnaissance automatisée des caractéristiques humaines dans les espaces accessibles au public devraient être interdites.

Viennent ensuite les systèmes d'IA « à haut risque », c'est-à-dire ceux qui présentent un risque élevé pour la santé, la sécurité ou les droits fondamentaux des personnes physiques. Il s'agit notamment des technologies d'IA utilisées dans les infrastructures

critiques (ex. transport), l'éducation ou la formation professionnelle, les services privés et publics essentiels (ex. évaluation du risque de crédit), les domaines du maintien de l'ordre, la gestion de la migration, de l'asile et des contrôles aux frontières ou encore l'administration de la justice, de même que les technologies d'IA utilisées dans les composants de sécurité des produits (ex. jouets, dispositifs médicaux, véhicules...).

« *Les techniques d'IA de notation sociale ou de reconnaissance automatisée des caractéristiques humaines dans les espaces accessibles au public devraient être interdites.* »

Ces systèmes d'IA à haut risque peuvent être mis sur le marché sous réserve du respect de certaines obligations et d'une évaluation préalable de la conformité qui pourra être fondée, selon le cas, soit sur un contrôle interne, soit sur une certification par un organisme notifié devant respecter des obligations d'indépendance, de compétence et d'absence de conflit d'intérêts. Parmi les obligations des systèmes d'IA à haut risque, on notera (i) l'établissement d'une



© Istock

documentation technique démontrant que le système satisfait aux exigences du règlement (ii) la fourniture d'informations aux utilisateurs sur les caractéristiques, capacités et limites de performance du système, les mesures de contrôle humain mises en oeuvre et la durée de vie attendue du système (iii) ainsi que la mise en oeuvre et le maintien à jour d'un système de gestion des risques.

Ce système de gestion des risques est comparable aux analyses d'impact menées par les responsables de traitement pour les traitements de données à caractère personnel susceptibles d'engendrer un risque élevé pour les droits et libertés des personnes physiques. Il consiste en effet à identifier et évaluer les risques connus et prévisibles du système et à adopter des mesures appropriées de gestion des risques. À cet égard, l'EDPB et l'EDPS estiment que la classification du système d'IA à risque élevé devrait déclencher une présomption de risque élevé au titre du RGPD, dans la mesure où des données à caractère personnel sont traitées.

La proposition de règlement prévoit aussi que certains systèmes d'IA présentant des risques spécifiques de manipulation (ex. deepfake audio ou vidéo) sont soumis à des obligations de transparence (ex. pour préciser que les contenus ont été générés ou manipulés artificiellement).

Enfin, la proposition de règlement prévoit la possibilité pour les fournisseurs de systèmes d'IA ne présentant pas de risques élevés de définir des codes de conduite destinés à appliquer volontairement des exigences obligatoires pour les systèmes d'IA à haut risque.

En termes de sanctions, les violations du règlement pourront donner lieu à de très lourdes amendes administratives (jusqu'à 30 millions d'euros ou 6 % du chiffre d'affaires annuel mondial), un niveau de sanction encore plus élevé que celui prévu par le RGPD. Or, l'EDPB et EDPS ont souligné, dans leur avis conjoint, l'enjeu majeur d'articulation de ce texte avec le RGPD, la classification d'un système d'IA comme étant à haut risque ne signifiant pas nécessairement qu'il soit licite en soi et qu'il puisse être déployé par l'utilisateur en tant que tel. Ce dernier sera dans la plupart des cas responsable des traitements de données à caractère personnel mis en oeuvre, tenu de

respecter les obligations prévues par le RGPD avant la mise en oeuvre d'un système d'IA appelé à exploiter des données personnelles.

À PROPOS DES AUTEURS



Christiane Féral-Schuhl,

Avocat associé du cabinet Féral-Schuhl / Sainte-Marie

Christiane Féral-Schuhl est avocat associé du cabinet Féral-Schuhl / Sainte-Marie. Elle est également auteur de l'ouvrage de référence *Cyberdroit*, paru chez Dalloz. Depuis plus de 30 ans, elle exerce ainsi dans le secteur du droit de l'informatique et des nouvelles technologies. Elle est également médiatrice agréée auprès du Centre de Médiation et d'Arbitrage de Paris (CMAP).

Elle a été bâtonnier du Barreau de Paris (mandat exercé de 2011 à 2013) et présidente du Conseil National des Barreaux (2018-2020).

Elle a participé à de nombreux groupes de travail et commissions officiels : membre de la commission parlementaire de réflexion et de propositions ad hoc sur « le droit et les libertés à l'âge du numérique » (rapport « Numérique et libertés : un nouvel âge démocratique » en 2015), personnalité qualifiée au Haut Conseil à l'égalité entre les femmes et les hommes (2013-2015).



Justine Sinibaldi

Avocate Senior Counsel et DPO au cabinet Féral-Schuhl/Sainte-Marie

Elle a co-rédigé cette tribune avec Christiane Féral-Schuhl.

UN ARTICLE RÉDIGÉ PAR

Christiane Féral-Schuhl et Justine Sinibaldi

/ PROJETS

Comment La Banque Postale se prépare à rendre les algorithmes explicables

Lors de l'événement DataOps.Rocks, organisé par Saagie le 23 septembre 2021, les data scientists de La Banque Postale ont expliqué comment l'entreprise se prépare dans la perspective de l'IA Act, en travaillant sur l'explicabilité des algorithmes.

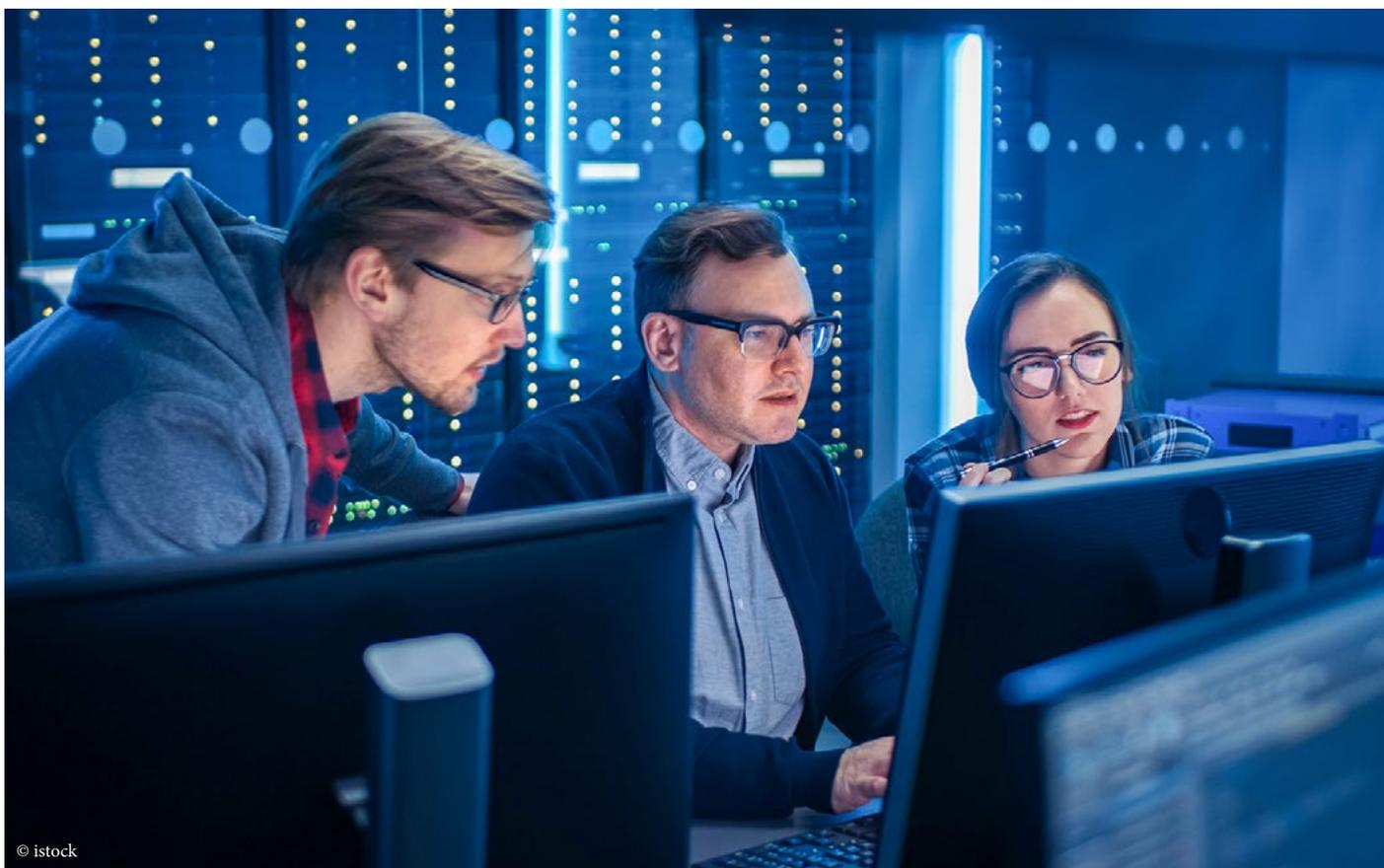


© DR-La Banque Postale

Clémence Panet-Amaro, chief data scientist à La Banque Postale : « Le projet d'IA Act est un pas vers une intelligence artificielle digne de confiance. »

Créée en 2006, La Banque Postale réunit les différents services financiers du groupe La Poste. Dans son plan stratégique 2020-2030, les données occupent une place essentielle, irriguant l'ensemble des axes. Pour répondre aux enjeux actuels autour de la donnée et se préparer à ceux à venir, La Banque Postale a créé en 2021 une nouvelle direction de l'innovation, du digital et de la data. Au sein de cette direction, la data factory, confiée à Clémence Panet-Amaro, travaille notamment sur l'explicabilité des algorithmes, un sujet amené à prendre de l'ampleur avec le projet de réglementation IA Act, soumis au Parlement européen et au Conseil de l'Union européenne le 21 avril 2021. Lors de l'événement DataOps.Rocks, organisé par Saagie le 23 septembre 2021, Clémence Panet-Amaro, chief data scientist à La Banque Postale, Dilia Carolina Olivo, data scientist à La Banque Postale et Bilal Azenoud, data scientist chez Probayes, ont présenté leurs travaux.

« La data factory fonctionne un peu comme un cabinet de conseil interne, avec huit data scientists qui accompagnent la banque et ses filiales sur tous leurs projets data, animent et fédèrent la communauté data », a expliqué Clémence Panet-Amaro afin de préciser le rôle de son équipe. Les projets sur lesquels celle-ci intervient touchent de nombreux domaines : le digital, le marketing, la lutte contre la fraude et le blanchiment, la gestion du risque ou encore le juridique et la conformité. « Nous allons aussi sur certains



sujets un peu plus novateurs, comme l'amélioration de processus transverses », a indiqué la chief data scientist.

L'IA Act, un texte basé sur les impacts humains

Pour une banque, amenée à traiter de nombreuses données clients sensibles, il est essentiel de garder le lien de confiance avec les clients. Dès 2016, le groupe La Poste a mis en place une charte data, posant des principes de transparence et d'usage raisonné des données. Lors du rapprochement avec la Caisse des Dépôts et CNP Assurances en 2020, l'ensemble des chartes existant dans le groupe ont été partagées, afin de s'assurer de la cohérence globale des principes posés. Sept piliers guident aujourd'hui la stratégie data du groupe : la robustesse technique et la sécurité ; le respect de la vie privée et la gouvernance des données ; la transparence ; la diversité, la non-discrimination et l'équité ; le bien-être sociétal et environnemental ; la responsabilité et enfin une action et un contrôle humains. La capacité à rendre les modèles basés sur la donnée explicables, en particulier ceux d'intelligence artificielle, contribue directement à plusieurs de ces

principes. Toutefois, en démarrant leur travail sur l'explicabilité, les data scientists de la data factory ont rencontré de nombreuses questions : quels modèles faut-il expliquer ? Seulement les modèles de machine learning ou bien les modèles déterministes ? À qui ? À celui qui les conçoit, à celui qui les contrôle, à celui qui est visé par la décision de l'algorithme ? Quel est le niveau de technicité requis et comment expliquer les modèles ?

« Le projet d'IA Act est un pas vers une intelligence artificielle digne de confiance »

Pour Clémence Panet-Amaro, « le projet d'IA Act est un pas vers une intelligence artificielle digne de confiance », qui réduit le champ des possibles et répond aux questions précédemment énumérées. Le projet se fonde sur une approche par les risques et sur l'impact humain pour classer les modèles en quatre catégories, de la plus risquée à la moins risquée. Le niveau 1 désigne des systèmes jugés inacceptables, comme ceux qui manipulent le comportement humain. Dans le niveau 2, très risqué, figurent des processus

comme l'octroi de crédit ou le recrutement. Le niveau 3 concerne des systèmes autorisés, mais soumis à des obligations de transparence (chatbots, contenus modifiés par IA), tandis que le niveau 4, avec un risque minime ou nul, recouvre par exemple des processus comme la maintenance prédictive. La proposition de loi s'applique à presque à tous : fournisseurs, utilisateurs, importateurs et distributeurs de systèmes d'IA au sein de l'Union européenne. Enfin, tous les modèles sont visés, même les classiques modèles experts et de statistique. Les sanctions encourues sont assez substantielles et cumulatives : par exemple, 30 millions d'euros ou 6 % du chiffre d'affaires annuel en cas de non-respect des usages interdits. « Elles sont prévues pour être très dissuasives », a souligné Clémence Panet-Amaro.

Identifier les variables qui ont pesé dans une décision

La Banque Postale estime que le texte de l'IA Act pourrait entrer en application en 2024. C'est notamment dans cette perspective que la data factory travaille sur l'explicabilité, même si d'autres régulations, parmi les multiples textes régissant le secteur bancaire, ont déjà alimenté la réflexion en amont. Les travaux sont menés conjointement avec la maison-mère La Poste, Probayes, (filiale du groupe La Poste depuis 2016, spécialisée depuis sa création dans les usages de l'intelligence artificielle) et l'école Telecom Paris. « Nous voulons construire une boîte à outils d'explicabilité et de lutte contre les biais, partagée et intégrée par les différentes entités du groupe », a indiqué Clémence Panet-Amaro.

« *Nous voulons construire une boîte à outils d'explicabilité et de lutte contre les biais, partagée et intégrée par les différentes entités du groupe* »

Les trois data scientists ont ensuite présenté quelques-unes des méthodes utilisées pour rendre les modèles explicables. Pour les modèles prédictifs, il existe deux niveaux d'explication, un niveau global

et un niveau local. Au niveau global, il s'agit d'identifier les différentes variables de données examinées par le modèle et de mesurer leur importance respective, tandis qu'au niveau local, le but est d'expliquer quelles variables ont pesé dans une décision. « Il s'agit par exemple d'expliquer à une personne pourquoi elle n'a pas obtenu son crédit. Ce deuxième niveau est important pour le client et le conseiller », a souligné Clémence Panet-Amaro. En termes d'outils, les équipes utilisent des approches assez usuelles, comme la perturbation (SHAP, LIME) et les méthodes contrefactuelles (ICE, DICE). Pour les modèles basés sur de la reconnaissance d'image, les data scientists se servent de méthodes comme l'occlusion, le gradient descent et les integrated gradients afin d'identifier les zones de l'image qui influencent la classification. Dilia Carolina Olivo a ensuite illustré la méthode d'occlusion avec un exemple portant sur la reconnaissance faciale. « La méthode consiste à cacher de manière itérative une partie de l'image, pour analyser les différences avec et sans les pixels cachés. L'objectif est de trouver la zone du visage qui contribue le plus à une classification positive (c'est la même personne) ou négative (ce n'est pas la même personne) », a-t-elle expliqué.

À PROPOS DE LA BANQUE POSTALE

Créée en 2006, La Banque Postale réunit l'ensemble des services financiers du groupe La Poste : banque de détail, banque de financement et d'investissement, assurance vie et non-vie. La banque s'appuie sur un réseau de 17 000 points de contact, incluant les bureaux de poste, agences communales et commerces relais. Parmi ses filiales elle compte notamment la banque mobile My French Bank, la banque privée BPE et depuis 2020 CNP Assurances. Son produit net bancaire s'élève à 7,7 milliards d'euros et elle totalise 20 millions de clients.



UN ARTICLE RÉDIGÉ PAR

Aurélie Chandeze, Rédactrice-en-chef adjointe

/ PROJETS

AG2R La Mondiale s'appuie sur l'IA pour accélérer le traitement des mails de ses clients

AG2R La Mondiale utilise la solution de traitement automatique de documents de reciTAL, basée sur des technologies d'intelligence artificielle, pour accélérer les temps de réponse aux mails entrants dans ses centres de relation client.

© istock



© DR-AG2R

Cyril Durupt, directeur de projets smart automation chez AG2R La Mondiale :
« Le gain de productivité se fait grâce à l'IA, mais aussi par sa prise en compte dans le processus global. »

Spécialisé dans la protection sociale et patrimoniale, AG2R la Mondiale accompagne plus de 15 millions d'assurés et 500 000 entreprises sur l'ensemble du territoire français. Dans le cadre de ses activités, le groupe possède plusieurs centres de relation client (CRC) spécialisés, qui traitent les demandes des clients arrivant sur les différents canaux, notamment mail et courrier. Afin d'accélérer les délais de réponse aux clients, le groupe a décidé en 2019 de mettre en place un outil de traitement automatisé de documents, retenant la solution de reciTAL. Pour CIO, Cyril Durupt, directeur de projets smart automation chez AG2R La Mondiale, est revenu sur ce projet.

« Il y a quelques années, nous avons démarré une réflexion sur l'automatisation intelligente. Nous avons réalisé une veille au sein du groupe, pour identifier des innovations technologiques permettant d'améliorer les processus au sens large », relate Cyril Durupt. Dans ce contexte, un besoin a émergé dans les centres de relation client et de gestion, pour détecter plus efficacement les réclamations dans les flux entrants, qu'il s'agisse de mails ou de courriers papier préalablement numérisés. « Rapidement, nous avons compris que le périmètre de cette demande était plus étendu, et qu'il s'agissait davantage d'un besoin de classification global sur nos flux entrants », explique Cyril Durupt. Son équipe a alors sollicité les différentes directions concernées par ce besoin, en vue de lancer une phase



d'expérimentation sur les solutions possibles. Parmi celles-ci figuraient notamment les différents CRC organisés par métiers (santé, prévoyance, épargne, retraite), mais aussi les directions de gestion collective et individuelle, ainsi que des directions internes, comme la direction des ressources humaines. « Avec 14 000 collaborateurs, les équipes RH reçoivent beaucoup de demandes à traiter », note le directeur de projets. La direction de la logistique, chargée de l'acquisition des flux entrants, notamment les courriers qui doivent être qualifiés et numérisés, était aussi impliquée dans la démarche.

Automatiser le routage des mails

Une fois que les différentes parties prenantes ont donné leur accord pour l'expérimentation, l'équipe a réalisé un benchmark de toutes les solutions pouvant répondre au besoin. « Nous avons rencontré des start-ups, travaillé avec notre direction de l'innovation, avec la direction des systèmes d'information et de l'organisation (DSIO) et les métiers pour identifier des offres pertinentes », indique Cyril Durupt. Le benchmark s'est peu à peu structuré autour de deux axes : d'un côté, des acteurs historiques, avec des technologies de

classification de type moteur de règles ; de l'autre des entreprises plus récentes, proposant des systèmes de classification basés sur des technologies d'intelligence artificielle (IA). « Pour l'expérimentation, nous avons finalement retenu deux outils basés sur l'IA, celui de reciTAL et une offre de CGI », indique le directeur de projets. En septembre 2019, AG2R La Mondiale a lancé une double expérimentation, avec les mêmes jeux de test auprès des deux acteurs. « Nous avons dû construire un data set pertinent pour l'apprentissage machine. Les humains ont un rôle important lors de cette étape : en effet, les résultats finaux dépendent de la capacité à classer correctement les données pour créer le modèle qui servira ensuite à entraîner l'IA », souligne Cyril Durupt. L'expérimentation proprement dite a ensuite duré entre 3 et 4 mois, avec un critère clair pour déterminer le succès : « Nous voulions que le taux d'erreurs de classification soit moindre que celui d'un collaborateur humain », affirme Cyril Durupt. Les résultats se sont révélés concluants pour les deux solutions, validant la technologie d'IA. « Nous n'avons pas perçu la pertinence de solutions hybrides, couplant IA et moteur de règles. Une ou deux règles pourraient faciliter les choses dans certains cas précis. En revanche, avoir trop de règles déterministes risque

de perturber l'IA », observe Cyril Durupt. Le choix s'est finalement porté sur l'outil de reciTAL, qui se distinguait notamment sur la détection des réclamations.

Fin mars 2020, la phase expérimentale achevée, le nouvel objectif était de mettre en oeuvre la solution dans le cadre d'un projet. Pour cela, il fallait valider le déploiement d'un point de vue métier et économique, avec une analyse de rentabilité. Afin de justifier l'investissement, l'équipe a bâti un modèle économique sur plusieurs années. Celui-ci mise principalement sur les gains de productivité pouvant être obtenus grâce à l'automatisation de tâches chronophages, comme l'ouverture, l'analyse et le routage des mails : l'équipe espère gagner jusqu'à 20% de temps par rapport à la situation initiale. Cette phase d'analyse a permis d'obtenir le feu vert six mois plus tard, en octobre 2020.



Afin de justifier l'investissement, l'équipe a bâti un modèle économique sur plusieurs années et espère gagner jusqu'à 20% de temps par rapport à la situation initiale.

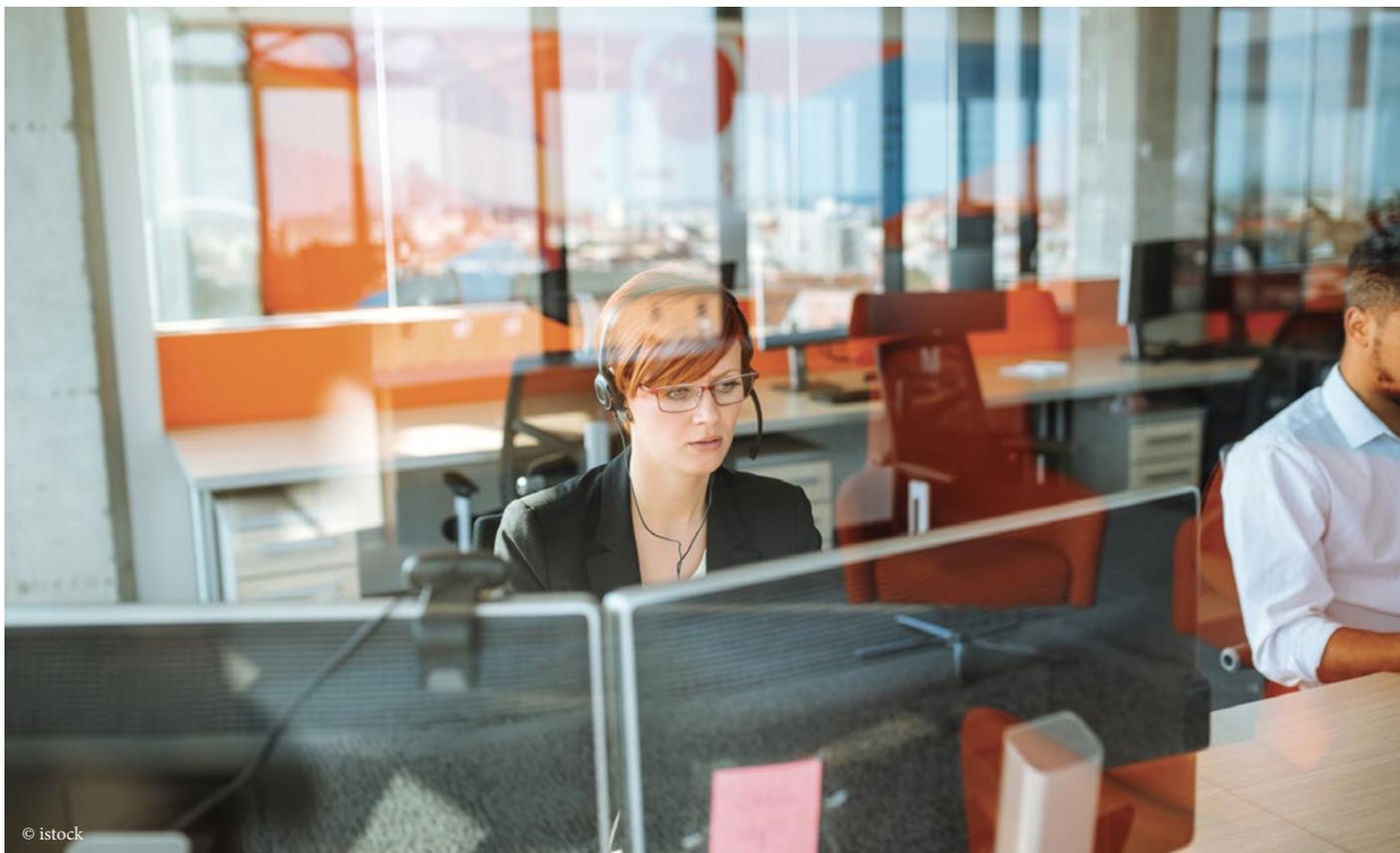
AG2R La Mondiale a alors entrepris de déployer reciTAL sur un premier périmètre, un CRC gérant les relations avec les entreprises, qui reçoit environ 110 000 mails par an. « Environ 45 % de ce flux ne peut être traité par les conseillers du CRC. L'action requise consiste alors à pousser les mails vers les bons services de gestion. Le modèle d'IA a permis d'automatiser correctement une grande partie de ce routage depuis sa mise en service en avril 2021 », souligne Cyril Durupt. Les mails et courriers entrants sont dirigés dans la solution de ReciTAL, qui les répartit en 13 catégories, avec plus de 90 % de précision. Selon la catégorie identifiée, les mails sont transférés automatiquement vers les centres de gestion ou traités par le CRC pour les réponses de niveau 1. À la clef, les gestionnaires du CRC ont déjà gagné plus de 15 % de temps sur le traitement des mails des clients, notamment sur le routage et la préanalyse. « Même sur les flux restant au niveau 1, il y a un gain grâce à la classification préalable, même s'il est plus difficile à mesurer », confie Cyril Durupt.

« La solution permet de mieux organiser le travail dans les CRC. Par exemple, en filtrant les mails catégorisés comme urgents, les responsables des centres peuvent plus facilement allouer une équipe à leur traitement pendant quelques jours. »

La qualité du data set, critère déterminant

Du côté des collaborateurs du CRC, l'outil a reçu un très bon accueil selon le directeur de projets. « Avec les responsables du centre, nous avons voulu faire en sorte que le projet ait le moins d'impact possible pour les conseillers. Nous fournissons une information claire sur la classification, et ensuite ils n'ont qu'une seule action à faire : cocher une case s'ils jugent la classification erronée. » Ces flux-là, identifiés comme des anomalies, remontent régulièrement dans le moteur de reciTAL. « La console d'administration fournie avec l'outil permet de vérifier que le taux de réussite de prédiction reste au niveau attendu. Nous avons la capacité de réentraîner l'IA nous même si besoin, pour améliorer ou maintenir le taux de reconnaissance, ou si la typologie des mails change de façon importante », indique Cyril Durupt. Le premier projet a suscité un élan positif au sein du groupe. En décembre 2021, un deuxième déploiement a eu lieu sur le CRC Santé d'AG2R La Mondiale, avec un périmètre trois fois plus gros que le précédent : environ 300 000 mails traités par an. « Nous avons choisi de cibler nos efforts là où les volumes sont importants », précise Cyril Durupt.

À présent, l'équipe échange avec les directions de gestion collective et individuelle pour identifier d'autres opportunités. Elle travaille également sur l'enrichissement et l'optimisation du dispositif afin d'augmenter encore les gains d'efficacité. « Nous savons que nous n'aurons jamais un taux de reconnaissance de 100 %. Pour réaliser les gains souhaités, il faut prendre le processus dans sa globalité et déployer tout ce que nous avons imaginé autour », explique Cyril Durupt. Il s'agit par exemple d'outils pour accueillir les mails routés de façon automatique, ou de fonctionnalités de réponse semi-automatique aux mails, couplées à la classification pour accroître le gain de temps dans les CRC. « Le gain de productivité se fait grâce à l'IA, mais pas par l'IA seule, c'est aussi sa prise en compte dans le



processus global qui détermine les résultats obtenus », insiste Cyril Durupt. « Il faut considérer le processus dans son ensemble, pour bien comprendre où on va placer la solution et quel impact elle aura autour. » Parmi les facteurs qui contribuent au succès dans de tels projets, Cyril Durupt évoque aussi la qualité du data set initial, très déterminante. Son équipe a travaillé avec les data scientists de reciTAL afin de s'en assurer, en veillant notamment aux effets de saisonnalité. « Certains types de mails arrivent davantage en début ou en fin d'année », explique-t-il. Le directeur de projets smart automation estime également qu'il faut prendre le temps nécessaire pour bien comprendre la typologie des flux à traiter. « Nous avons constaté que cela fonctionne bien avec des classifications comptant entre 12-15 catégories. Au-delà de 20, c'est un peu plus compliqué. » Pour finir, Cyril Durupt mentionne aussi l'accompagnement fourni par reciTAL. « Leurs équipes nous ont aidés sur l'acculturation lors de la phase expérimentale. Elles se sont également montrées réactives lors de la mise en production, pour traiter les problématiques d'intégration inhérentes à de tels projets. » La DSIO du groupe a elle aussi apporté son appui, notamment lors du déploiement. « Aujourd'hui, leurs collaborateurs sensibilisent les autres équipes informatiques du groupe, contribuant eux aussi à la

vulgarisation de l'IA au sein d'AG2R La Mondiale », souligne Cyril Durupt. Et les perspectives s'annoncent nombreuses, avec de nombreux usages encore à explorer, par exemple sur les traitements documentaires.

À PROPOS D'AG2R LA MONDIALE

Spécialiste de la protection sociale et patrimoniale, AG2R La Mondiale est une société de personnes à gouvernance paritaire et mutualiste. Le groupe est au service de plus de 15 millions d'assurés et 500 000 entreprises. Présent sur l'ensemble du territoire métropolitain et ultramarin, AG2R La Mondiale emploie plus de 14 000 collaborateurs.



UN ARTICLE RÉDIGÉ PAR

Aurélie Chandeze, Rédactrice-en-chef adjointe