© greenbutterfly

DÉFINIR LE CADRE NORMATIF D'UNE IA DE CONFIANCE DANS LES ENTREPRISES

MÉDÉRIC CHOMEL (X00)
directeur data IA automatisation d'Orange France

Les inquiétudes provoquées par le développement de l'intelligence artificielle (IA) amènent les pouvoirs publics à préparer des mesures de régulation, qui ne seront en vigueur que dans quelques années. Les entreprises ont tout intérêt à précéder la mise en place de ces normes, en réfléchissant à une autorégulation qui les place en position dynamique et non en sujet de contrôle.

Au cours de ces dernières années, les entreprises ont accéléré fortement sur l'utilisation de l'intelligence artificielle, afin de répondre à leurs enjeux stratégiques. Dans le même temps, nous observons une défiance croissante envers ces nouvelles technologies de la part des clients et des collaborateurs de nos entreprises (selon un sondage Ipsos de janvier 2022, près de 40 % des Français souhaitent une IA de confiance). Face à ce constat, un double mouvement

s'amorce pour promouvoir une IA de confiance, avec d'un côté des travaux de régulation et de normalisation, mais également de l'autre côté des initiatives volontaires de la part des entreprises. Commençons par nous accorder sur une définition de l'IA de confiance. Aujourd'hui, la plupart des initiatives sur le sujet partent de la définition proposée par la Commission européenne et des sept exigences auxquelles les systèmes d'IA doivent répondre. Un système d'IA de confiance est donc un système qui a évalué et mitigé le risque de dérive sur chacun des sept piliers qui suivent.

Des contrôles humains sur le traitement

Nous retrouvons ici la notion de prise de décision assistée par l'IA et non prise par une IA. Par exemple, dans les systèmes de recommandations d'offres qu'Orange met à disposition de ses vendeurs (proposition commerciale personnalisée, affichée dans les outils des vendeurs), c'est toujours le vendeur qui décide quelle offre proposer au client (sans obligation de proposer l'offre qui est dans le système).

Robustesse technique et sécurité du traitement

Lorsque des algorithmes doivent agir sur des systèmes avec un fort enjeu sécuritaire, la robustesse est clé dans les opérations afin d'éviter des dérives tout au long du cycle de vie. Par exemple, lorsque les systèmes d'IA d'Orange traitent des images de poteaux téléphoniques qui penchent pour prédire le risque de chute, ces systèmes doivent être fiables à chaque instant.

Respect de la vie privée et gouvernance des données

Nous retrouvons ici tout le domaine ouvert par le RGPD depuis quelques années (collecte des données nécessitant souvent un consentement, traitement pour une finalité définie, droit à l'oubli, etc.).

Impact environnemental et sociétal

À ce jour, l'impact des systèmes d'IA sur l'environnement est assez faible (de l'ordre du pourcent des émissions globales de CO_2 selon différentes études, chiffre complexe à calculer) mais en croissance forte et avec de vraies solutions pour agir. Optimiser les données avec lesquelles nous entraînons les modèles, choisir les bonnes heures pour les faire tourner et avoir ainsi l'énergie la plus verte (solaire et éolienne notamment), éviter de surentraîner les modèles, choisir la localisation des serveurs avec l'énergie la plus verte peuvent être des solutions vertueuses et facilement applicables dans les projets d'IA.

Diversité, non-discrimination et équité

C'est le sujet le plus repris dans les médias quand on parle d'éthique de l'IA. La question centrale est de savoir comment éviter que nos systèmes d'IA créent ou renforcent des biais que nous ne souhaitons pas accepter dans une situation donnée. Les exemples les plus connus concernent les biais de genre pour le recrutement, les biais d'origine ethnique pour les demandes de prêt, etc. En juin 2022, Facebook a été sanctionné par la justice américaine car son algorithme Lookalike Audience a été jugé discriminatoire, en ce sens qu'il ne présentait des offres de location qu'à certains utilisateurs (en fonction de leur religion, l'origine ethnique, le sexe, le statut marital…).

Transparence autour du traitement

Un des problèmes majeurs des systèmes d'IA modernes est qu'ils reposent sur des algorithmes dits « boîte noire », dans lesquels nous n'avons pas d'explication simple pour chaque résultat. Par exemple, pourquoi un algorithme de prévision d'appels en service client, qui va prédire une certaine valeur, ne vient pas d'une équation mathématique simple et explicable mais d'un ensemble très complexe de traitements ? Un des sujets majeurs actuellement sur l'IA est donc soit de recourir plus fréquemment à des modèles explicables, soit de mettre en place des solutions pour mieux expliquer les résultats des algorithmes à ceux qui les utilisent, en particulier. →

→ Les responsabilités associées au traitement

Tout traitement automatisé engage la responsabilité d'un acteur en cas d'erreur ou de dérive. Ainsi les premiers véhicules autonomes nécessitaient la présence d'un humain au poste de conduite, qui était alors responsable et, en cas de problème du véhicule autonome, pouvait à tout moment reprendre la maîtrise. Aujourd'hui, l'identification et le partage de la responsabilité entre les différentes parties prenantes d'un projet restent impératifs, notamment par le biais contractuel, pour assurer une maîtrise de la chaîne de responsabilités et pour permettre aux utilisateurs et personnes concernées de se retourner vers les bons interlocuteurs, en cas de dérive ou d'accident lié à un système d'IA.

Des projets de régulation et de normalisation attendus en 2025

La Commission européenne travaille depuis quelques années sur un *AI Act* qui, dans la lignée du RGPD mis en place en 2018, vise à réguler l'utilisation de l'IA. Pour rendre sa régulation opérante, elle a eu l'excellente idée de proposer une approche par le risque. En effet, tout système d'IA ne nécessite pas une analyse détaillée de chacun des sept piliers et l'effort des producteurs d'IA doit donc se focaliser sur les systèmes les plus à risque. Une première liste de systèmes d'IA dits « critiques » a même été établie (recrutement, santé, éducation, etc.). C'est donc une régulation dont l'objectif principal est d'arriver à établir des principes d'action permettant à la fois d'atteindre une IA de confiance et de permettre l'innovation, en focalisant les efforts sur les cas les plus à risque.

En parallèle de la finalisation de la rédaction de l'*AI Act*, de nombreux organismes de normalisations travaillent de concert avec des acteurs privés pour une mise à disposition progressive de normes sur l'IA. Parmi plus d'une trentaine de normes sur l'IA publiées ou en cours de rédaction, nous pouvons par exemple citer la norme ISO/IEC 24027, biais des systèmes d'IA (déjà publiée) ou la norme ISO/IEC 42001 : AI Management System en cours d'écriture.

Des initiatives volontaires de la part des entreprises

Au-delà de ces travaux de régulation, nous voyons également une approche englobante poussée par certaines entreprises, dont Orange, qui repose sur quatre grands piliers.

"La Commission européenne travaille sur un AI Act pour réguler l'utilisation de l'IA."

Des engagements responsables. Bien que les chartes soient trop souvent inopérantes lorsqu'elles restent théoriques, elles représentent cependant une condition nécessaire à la mise en œuvre d'une démarche exigeante, en ce sens que la charte ou les engagements de l'entreprise viennent définir, au-delà de ce que la loi nous dit, une direction dans laquelle aller. Prenons un exemple concret : chez Orange, nous avons décidé que, dans les cas de collaboration humain-IA, l'humain aurait toujours le dernier mot. Par exemple nous utilisons des algorithmes de reconnaissance visuelle pour aider nos techniciens sur le terrain mais, si le technicien ne veut pas suivre la recommandation de l'algorithme, il en a le droit sans aucun impact sur son évaluation.

Acculturation et culture d'entreprise. Pour que les équipes qui font mais aussi qui utilisent l'IA puissent construire une IA de confiance, il faut également s'assurer d'une compréhension globale du sujet, notamment en définissant ce qu'est un système d'IA et comment il fonctionne. Nous avons constaté deux grands écueils dans notre mise en œuvre : soit un manque de vigilance dans l'identification des risques (technophilie aveugle), soit un rejet permanent de l'innovation technologique (technophobie de principe). L'acculturation et l'accompagnement visent à éviter ces écueils, pour avoir une approche plus pragmatique.

Organisation et comités. Afin d'aborder sereinement ce sujet, il convient aussi de définir de nouveaux rôles ainsi qu'une gouvernance qui permettra d'évaluer les risques, aider à l'identification de solutions et émettre des recommandations.

Processus et outils. Les engagements responsables ne deviennent opérants qu'une fois qu'ils se retrouvent dans le code lui-même des systèmes d'IA. Nous devons donc aider nos *data scientists* à utiliser les bons outils pour rendre cela possible. Nous parlons ici de librairies *open source* ou d'outils spécifiques pour mesurer la dérive des algorithmes sur les biais par exemple.

Une question à traiter sans attendre

Il me semble essentiel que les entreprises s'emparent dès maintenant de ce sujet sans attendre la régulation qui arrivera au minimum dans deux ans. En effet, c'est une transformation en profondeur des modes de gestion de projets d'IA qu'il vaut mieux mettre en place dès le début, plutôt qu'en rattrapage. De plus, cela nous permettra d'en tirer un vrai bénéfice (interne et externe), plutôt que de la subir comme cela a été le cas avec le RGPD. ×