

Comment concilier innovation et éthique ?

Description

L'idée d'une IA universellement "juste" est utopique. Cela ne signifie pas pour autant que la question doit être évacuée, au contraire.

Le 11 février 2025, à l'issue du sommet pour l'action sur l'intelligence artificielle; à Paris, l'Élysée publiait une déclaration signée par une soixantaine de pays afin de promouvoir une IA inclusive et durable. Cette déclaration s'appuie sur un nombre croissant de travaux portant sur les principes à adopter pour favoriser une utilisation éthique de l'IA et notamment ceux produits par l'Unesco (2021), l'ONU (2022) et l'OCDE (2024). Si les déclarations de principes sont belles, que valent-elles face aux enjeux concurrentiels et à la course à l'innovation? On se souvient que de grandes entreprises américaines s'étaient dotées, dès 2018, de tels cadres et valeurs avant de se séparer leurs équipes en charge de l'éthique, ou de les neutraliser. C'était en 2023, quand la compétition pour le développement des IA génératives devenait la grande priorité.

Faudrait-il nécessairement choisir entre l'innovation et l'éthique? Cette dernière serait-elle la marque de ceux qui, arrivés trop tard ou n'ayant pas les ressources, se contenteraient d'être des suiveurs voire de simples utilisateurs? Non. Face à la tentation du déterminisme technologique, il est essentiel d'affirmer que les questions éthiques soulevées par le développement de l'IA sont majeures et donc à travailler. Au-delà des enjeux liés à l'empreinte environnementale des systèmes d'IA, la manière dont ceux-ci s'insèrent et transforment les pratiques et processus de travail se doit d'être réfléchie. Il est aussi nécessaire de s'interroger sur les évolutions, voire les ruptures, des modèles d'activité engendrés par le recours à l'IA dans les industries et les services: il en va de l'accessibilité des offres et de l'inclusion des personnes, mais également de la qualité des futurs économiques et sociétaux. Il convient, en conséquence, de s'interroger sur les tentatives échouées et de tenter de procéder différemment.

Un point commun existe entre déclarations de principes et affichages de valeurs qui constituent les cadres éthiques promus par les entreprises, les pouvoirs publics et autres organisations intergouvernementales. Celles-ci procèdent d'une approche classique, fixiste parce qu'essentiellement déontologique, de l'éthique: il s'agirait, pour se prémunir des dangers identifiés, de se donner un certain nombre de règles (théoriques) érigées en amont de l'expérience (pratique) de manière à limiter cette dernière. Or, deux problèmes majeurs accompagnent cette approche.

Premièrement, il est impossible de savoir à l'avance comment les technologies vont évoluer, quels nouveaux usages celles-ci vont engendrer ou comment l'acceptabilité sociale des technologies et des usages va évoluer. Ainsi, chercher à se doter de règles en amont de l'expérience conduit nécessairement à voir s'opposer à l'éthique, fixée à un temps t , l'innovation en train de se faire et son lot de questions et d'attentes nouvelles, logiquement non traitées au sein des cadres préexistants. Deuxièmement, l'approche classique ne débouche pas sur autre chose qu'un énoncé de valeurs générales (IA «juste», «de confiance», etc.) qui ne sont d'aucune aide pour guider l'innovation. Pour le dire dans le langage des pragmatistes, le sens des valeurs ne préexiste pas au jugement pratique. Lorsqu'une situation soulève un questionnement éthique, c'est précisément parce que les valeurs en vigueur se trouvent indéterminées face à la nouveauté de la situation. Ce sont, à l'inverse, les décisions prises en situation qui donnent progressivement sens à ce que les personnes et les

organisations entendent à travers les valeurs qu'ils professent.

Si l'approche classique de l'éthique de l'IA se révèle impuissante à accompagner le développement technologique en permettant de traiter les questionnements qui émaillent les processus d'innovation, une autre voie est cependant possible. Cette voie consiste à adopter une approche pragmatiste, c'est-à-dire en mettant en œuvre un questionnement partant des cas d'usage et en considérant que les décisions sont susceptibles d'évoluer au cours du temps, que celles-ci doivent résulter d'une délibération pluraliste engageant les différentes parties prenantes de l'entreprise (selon les cas les développeurs des systèmes d'IA, les métiers, les collaborateurs, les dirigeants, les partenaires, etc.) et, enfin, que la responsabilité d'identifier et soulever les doutes éthiques appartient avant tout aux équipes engagées dans les processus d'innovation.

Ce renversement d'approche est inévitable. Prenons l'exemple de l'équité algorithmique. Les biais qui peuvent exister dans un système d'IA ne sont pas toujours «objectifs», c'est-à-dire l'effet de problèmes méthodologiques rencontrés au cours du développement (un biais de sélection, par exemple). À vrai dire, les biais sont majoritairement «subjectifs», construits, relatifs aux modèles de justice ayant cours dans un espace de temps et de lieu donné. Nous reconnaissons une décision ou un comportement comme étant biaisés parce que nous appliquons à leur endroit une théorie de la justice particulière. Une réalité statistique peut-elle être biaisée? Tout dépend de la théorie de la justice à travers laquelle on la regarde et il n'est nul besoin d'IA pour le comprendre.

À titre d'exemple, en 2011, la Cour de justice de l'Union européenne a interdit aux assureurs de pratiquer des prix différenciés selon le sexe de leurs assurés, arguant que de telles pratiques étaient discriminatoires. Jusqu'alors, les jeunes conducteurs devaient s'acquitter d'une prime plus élevée que les jeunes conductrices, en lien avec le coût des sinistres les concernant. Selon la théorie de la justice empruntée pour considérer cette décision, on peut voir en elle l'affirmation d'une équité entre les sexes ou bien, au contraire, d'une iniquité devant la réalité des statistiques. Il n'y a jamais, en cette matière, de réponse définitive et les algorithmiques n'échappent pas à cette vérité: il n'est pas possible pour un algorithme d'être intrinsèquement juste car «la» justice n'existe pas. Il existe seulement des conceptions particulières de ce qui est juste, équitable, responsable, solidaire, etc. qui doivent être convoquées pour débattre de l'implémentation et du suivi d'un système d'IA.

Considère-t-on que les données d'entraînement sont biaisées et que le système d'IA, en les incorporant, l'est également? Considère-t-on que les comportements de certains utilisateurs du système sont biaisés et que ce dernier, par des mécanismes de renforcement, les incorpore? Ces questions appellent à expliciter et faire dialoguer dans le temps des expériences et intelligences plurielles afin de construire une position d'entreprise à la fois responsable et évolutive. Bien d'autres questions, dans le développement et le suivi de l'utilisation des systèmes d'IA, nécessitent une telle approche pragmatiste partant des cas d'usage. Le besoin d'explicabilité d'un système, par exemple, dépend de la criticité de celui-ci dans la chaîne de valeur. Selon qu'une entreprise doit pouvoir justifier le traitement des données opéré et les décisions rendues par un système qu'elle utilise, les technologies à employer pour le construire sont à débattre. Dans chaque cas, ce sont des réponses singulières et évolutives qui doivent être apportées. Abandonner l'éthique au prétexte d'une course à l'innovation ne peut désormais qu'être vu que comme une forme d'irresponsabilité.

Après des études de philosophie puis d'économie, Jonathan Sambugaro, auteur d'une thèse sur l'élaboration de la stratégie dans les grandes entreprises de l'économie sociale et solidaire, est maître de conférences à l'IAE Lille et co-dirige le Master Management des Organisations de l'ESS. Il est membre des conseils d'administration de la Macif et d'Aéma Groupe.

Categorie

1. Dernier numéro

Tags

1. BM35

date créée

mai 2025

Auteur

williame