

# Peut-on programmer le sens de l'éthique ?

Les algorithmes ne disposent pas de l'intuition humaine pour développer des valeurs comme l'équité. Cette situation place les programmeurs de systèmes d'intelligence artificielle face à des dilemmes éthiques. *Markus Christen*

**Abrégé** Les succès impressionnants de l'intelligence artificielle (IA) ont récemment donné lieu à moult débats éthiques. En moins d'un an, de nombreuses recommandations internationales sur l'IA « éthique » ou « digne de confiance » ont été publiées. Une étude de la Fondation pour l'évaluation des choix technologiques (TA-Swiss) à paraître début 2020 examine en profondeur les questions éthiques qui se posent dans différents domaines d'application de l'IA. Des craintes découlent surtout de certaines caractéristiques techniques des nouvelles formes d'IA, notamment leur manque de transparence. L'un des principaux défis vient du fait que l'automatisation de décisions à dimension éthique ne peut se faire sans définir de règles dans ce domaine. Qui doit les établir ? Est-ce une menace pour la liberté des humains ?

**O**n pourrait croire à un concours entre organisations internationales : en moins d'un an, l'Organisation de coopération et de développement économiques (OCDE) a publié des lignes directrices en matière d'intelligence artificielle (IA) et de protection des données, le Conseil de

l'Europe une « Recommandation sur la convergence technologique, l'intelligence artificielle et les droits de l'homme » et le groupe d'experts de haut niveau de l'Union européenne des « Lignes directrices en matière d'éthique pour une IA digne de confiance ». La liste des souhaits ainsi adressés par les experts en éthique aux développeurs d'IA est longue.

Un bon exemple est également donné par le rapport du Groupe européen d'éthique des sciences et des nouvelles technologies qui conseille la Commission européenne. Selon ce texte, l'IA ne doit pas porter atteinte à la dignité humaine et l'autonomie des êtres humains doit être préservée. La recherche et les applications de l'IA doivent être responsables. L'IA doit par ailleurs contribuer à la justice et à l'égalité d'accès aux avantages qu'elle apporte. En

Nous nous fions à notre intuition pour prendre des décisions éthiques, ce que ne peuvent pas faire les ordinateurs.



SHUTTERSTOCK

matière de réglementation, les décisions clés sur l'application de l'IA doivent être prises de façon démocratique. L'état de droit et l'obligation de rendre des comptes doivent en outre être garantis. L'IA ne doit par ailleurs porter atteinte ni à la sécurité, ni à l'intégrité physique et mentale, ni à la protection des données, ni à la sphère privée. Enfin, elle doit être compatible avec la protection de l'environnement et le principe de durabilité.

Vaste programme, donc. Un pareil éventail d'exigences éthiques trahit sans doute la grande incertitude provoquée par les derniers résultats impressionnants de l'IA.

Cette nouvelle technologie se retrouve dans toute une série de questions controversées liées à la transition numérique. Il est donc difficile d'aborder le sujet de façon ciblée. Son potentiel est en outre souvent exagéré dans le débat public, ce qui complique encore l'analyse.

Une étude commandée par la Fondation pour l'évaluation des choix technologiques (TA-Swiss) et dirigée par Markus Christen («Digital Society Initiative» de l'université de Zurich), Clemens Mader (Laboratoire fédéral d'essai des matériaux et de recherche, Empa) et Johann Cas (Académie autrichienne des sciences) s'est, dans ce contexte, penchée sur les défis sociétaux posés par l'IA<sup>1</sup>, en particulier sous l'angle de l'éthique.

## Une boîte noire

Sur le plan éthique, six caractéristiques techniques, économiques et sociales de l'IA doivent être considérées. Premièrement, l'IA ressemble de plus en plus à une boîte noire. Les nouvelles formes d'apprentissage automatique («machine learning») sont beaucoup moins transparentes pour les programmeurs d'algorithmes d'IA que «l'IA classique» des systèmes experts. L'entraînement avec d'énormes quantités de données produit des modèles sans lien physique ou logique apparent avec les phénomènes étudiés. D'un point de vue éthique, cela pose d'une part un problème de sécurité: il devient en effet plus difficile de vérifier si un système fonctionne toujours de la façon souhaitée. D'autre part, de tels systèmes pourraient être utilisés dans des situations ayant une dimension éthique sans qu'il soit possible de retracer la logique de leurs

décisions. Les débats en cours aux états-Unis sur un système de recommandations aux juges concernant la libération conditionnelle des prisonniers donnent un bon exemple des problèmes éthiques soulevés<sup>2</sup>.

Deuxièmement, les résultats de l'IA peuvent être faussés par les données utilisées. Ces dernières peuvent en effet contenir des éléments de partialité ou un biais qui influencent le comportement de l'algorithme. Un tel biais causé par un mauvais choix des données servant à «former» l'IA peut certes être perçu comme un problème d'ordre technique. Il sera toutefois difficile à identifier en raison de l'important volume de données concernées. La situation est plus ardue encore lorsque les données représentent fidèlement l'état des faits, mais que ces derniers sont le résultat d'un ordre social perçu comme éthiquement problématique. Un système d'évaluation automatique des candidatures d'Amazon pénalisait par exemple systématiquement les femmes parce que l'algorithme avait intégré le fait que cette entreprise appartenait à un secteur dominé par les hommes. Un système d'IA formaté de cette manière perpétuerait ainsi une situation indésirable et saperait les efforts déployés pour améliorer la représentation des femmes.

## Exactitude ou équité ?

Le troisième défi éthique posé par l'IA est celui de l'équité des algorithmes. Leurs paramètres sont définis par des programmeurs qui peuvent privilégier certaines valeurs ou certains intérêts par rapport à d'autres. Le fait que les normes éthiques doivent pouvoir être transposées en «langage machine» compréhensible par les ordinateurs complique également le problème. Des études ont notamment montré que les algorithmes ne pouvaient pas satisfaire deux exigences éthiques de légitimité équivalente (par exemple l'exactitude et l'équité) en même temps. Quatrièmement, les systèmes d'IA ont le potentiel de renforcer considérablement le mouvement d'automatisation de l'économie. La peur des pertes d'emplois influence ainsi largement le débat public sur l'IA. Certes, la mesure de ces pertes est controversée, le potentiel de création d'emplois n'est pas clair et de tels bouleversements économiques ont déjà eu lieu à plusieurs

<sup>1</sup> Description du projet disponible sous TA-Swiss.ch. L'étude sera publiée début 2020.

<sup>2</sup> Voir Loi et Christen (2019).

reprises. Ils ont cependant toujours été accompagnés de tensions sociales et de crises.

En cinquième lieu, les entreprises qui contrôlent de grandes quantités de données disposent d'un avantage concurrentiel, ce qui favorise la constitution d'oligopoles de données. Les grandes plateformes comme Alibaba, Facebook et Google sont souvent citées à ce titre. Bien connu dans le contexte de l'économie de l'Internet, l'idée que « le gagnant prend tout » ne pose pas uniquement des problèmes de droit de la concurrence. Lorsqu'une poignée de systèmes d'IA influencent de façon déterminante les décisions dans des domaines ayant une dimension éthique, on peut craindre une forme de standardisation et de domination de certaines valeurs.

Enfin, le risque existe que l'IA soit utilisée à des fins de surveillance massive ou de « grand coup de pouce » (« big nudging »), c'est-à-dire de très larges tentatives d'influencer les comportements, par exemple pour favoriser le respect de l'environnement. Cela pose d'une part la question de savoir comment assurer la légitimité des objectifs visés et des moyens employés et, d'autre part, celle de la mesure dans laquelle l'utilisation d'une IA développée au sein de sociétés ayant des normes sociales et des traditions démocratiques différentes peut poser des problèmes éthiques ou politiques. Les applications militaires de l'IA relèvent également de cette catégorie de préoccupations et les scientifiques mettent d'ores et déjà en garde contre une course aux armements dans ce domaine.

Chacun de ces points peut évidemment être largement développé. Leur défi éthique commun réside toutefois dans la nécessité de rendre les processus de prise de décision par les systèmes d'IA transparents pour les humains. La question de savoir comment garder le contrôle des systèmes d'AI complexes doit être clarifiée. En ce qui concerne les applications concrètes, il s'agit de définir des règles de conception de ces systèmes permettant de garantir qu'ils soient dignes de confiance.

## Qui décide ?

La crainte souvent évoquée d'une perte de contrôle totale ne semble pas fondée. L'utilisation de systèmes d'IA repose en effet largement

sur des décisions humaines, de la conception des systèmes – c'est-à-dire « l'assemblage » des différentes technologies nécessaires comme l'apprentissage automatique – au choix des données d'entraînement. Les humains choisissent les contextes dans lesquels ils s'appuient sur l'IA et les utilisateurs disposent également de nombreux leviers de contrôle sur les systèmes.

Le vrai problème est plutôt posé par la difficulté fondamentale de développer des IA capables d'assimiler des valeurs éthiques complexes comme l'équité. Alors que nous disposons tous d'une certaine liberté d'interprétation et de décision en matière d'éthique, les systèmes d'IA n'ont pas ces capacités d'intuition : les règles de décision doivent être expressément définies pour chaque situation concrète. Une expérience a été menée pour déterminer les décisions qui, du point de vue des participants, devaient être prises par un véhicule autonome piloté par une IA et confronté à des dilemmes<sup>3</sup>. Devait-il par exemple privilégier la protection des passagers ou celle des piétons ? Les réponses obtenues, parfois contradictoires, mettent en évidence des différences culturelles.

Ce constat donne lieu à une nouvelle série de questions. Qui doit prendre ce type de décisions ? De quelle manière ? Quelles sont les conséquences pour la liberté humaine, qui nous permet de faire parfois les « mauvais » choix ? Le défi éthique posé par l'IA vient du fait qu'elle nous oblige à examiner en détail les conséquences éthiques difficiles des décisions qui lui sont confiées. C'est à la fois un risque et une chance.

3 Awad (2018).



**Markus Christen**

Directeur de la Digital Society Initiative, directeur d'un groupe de recherche à l'Institut d'éthique biomédicale et d'histoire de la médecine de l'université de Zurich

## Bibliographie

Awad Edmond et al. (2018). « The Moral Machine experiment ». *Nature*, 563: 59–64.  
Loi Michele et Christen Markus (2019). « How to include ethics in machine learning research ». *ERCIM News*, 116.