

KNOWLEDGE.ESSEC.EDU

# ESSEC

# KNOWLEDGE

## REVIEW



**L'intelligence artificielle  
et la transformation  
technologique**

MAY 2021

# ÉDITORIAL



« De loin, le plus grand danger de l'intelligence artificielle est que les gens concluent trop tôt qu'ils la comprennent. »

– Eliezer Yudkowsky, fondateur de la Machine Learning Research Institute

La mention de l'intelligence artificielle pourrait évoquer des images de robots, de maisons intelligentes ou d'automatisation des tâches. Elle peut susciter l'intérêt, l'enthousiasme, même l'inquiétude. Quelle que soit votre réaction, il est peu probable qu'elle vous laisse indifférent : les systèmes d'intelligence artificielle touchent pratiquement tout le monde, et le monde de demain promet d'être passionnant.

Notre monde évolue rapidement, et nous avons la possibilité - la responsabilité - de participer à ce changement. Les technologies numériques et la science des données ont un impact sur notre façon d'interagir, d'apprendre et de travailler. Les problèmes modernes exigent des solutions modernes, et celles-ci soulèvent de nombreuses questions.

La signature de l'ESSEC est « Enlighten. Lead. Change », et nous pensons notre responsabilité est de préparer les futurs leaders aux défis auxquels nous sommes confrontés en tant que société — et aussi de considérer les défis comme des opportunités. Bien que le monde de demain comporte encore de nombreuses inconnues, nous croyons au pouvoir de la curiosité. C'est pourquoi, avec ce numéro spécial, nous souhaitons partager les connaissances des professeurs de l'ESSEC et mieux comprendre l'impact de l'intelligence artificielle et de la transformation numérique sur la société.

Ce numéro spécial de ESSEC Knowledge contribue à la nouvelle stratégie RISE de l'ESSEC, qui positionne notre école à l'interface de la science, du management et du changement sociétal. Un élément de la stratégie RISE est la création du Metalab for Data, Technology & Society, un laboratoire interdisciplinaire de recherche scientifique et d'innovation pédagogique. Avec cela, nous investissons dans la recherche qui étudie les méthodologies, les utilisations et les impacts de l'intelligence artificielle, en utilisant une approche interdisciplinaire, réfléchie et responsable.

Ce numéro spécial, qui met en lumière les influences très diverses des systèmes d'intelligence artificielle, présente des articles de professeurs en analyse, sciences de la décision, économétrie, économie, entrepreneuriat, finance, systèmes d'information, management, marketing, philosophie et statistiques. Les professeurs de l'ESSEC ont fourni leurs analyses d'experts sur des sujets allant de l'utilisation de l'intelligence artificielle dans les ressources humaines et le marketing, la propriété des données, l'impact de l'IA sur les entreprises et la prise de décision, aux dilemmes éthiques et aux discriminations potentielles que posent les systèmes de gouvernance et d'aide à la décision basés sur l'IA. Nous parlons de l'impact potentiel de l'IA et les nouvelles technologies sur le monde du travail, des ressources humaines, de l'analyse et du marketing à la gouvernance de l'information, en passant par la manière dont la recherche est menée et leur utilisation dans des initiatives de durabilité. Nous abordons également la manière dont l'IA pourrait avoir un impact sur notre vie quotidienne, en discutant de son utilisation dans les décisions de crédit, de l'interaction entre l'IA et le genre, et d'autres implications sociétales, y compris la question de la diffusion de fausses informations sur les réseaux sociaux. Ensemble, ces sujets donnent un aperçu de ce à quoi pourrait ressembler l'avenir alors que nous libérons la puissance de la technologie et des algorithmes — et apprenons également comment utiliser cette puissance pour le bien.

Les articles inclus dans ce numéro spécial donnent un aperçu de la transformation scientifique de notre monde, stimulant les discussions et les débats afin que nous puissions imaginer à quoi pourrait ressembler le prochain chapitre de l'humanité. Selon les mots de feu Okwui Enwezor, qui a récemment été le commissaire de la Biennale de Venise « All the World's Futures », nous nous demandons « comment saisir, rendre compréhensible, examiner et articuler correctement l'inquiétude actuelle de notre époque ? Comment pouvons-nous donner un sens au bouleversement actuel ? »

**Julia Smith**, Rédactrice-en-chef, ESSEC Knowledge  
Professeurs **Guillaume Chevillon** et **Julien Malaurent**,  
Co-directeurs académiques du Metalab

# SOMMAIRE



6

## L'INTELLIGENCE ARTIFICIELLE IMPRÈGNE L'ENSEMBLE DU CYCLE ANALYTIQUE

**Nicolas Prat**  
Département des Systèmes d'Information, Sciences de la Décision et Statistiques



10

## INTELLIGENCE ARTIFICIELLE ET RESSOURCES HUMAINES : ET SI ON JOUAIT À PILE OU FACE ?

**Valery Yakubovich**  
Département de Management



14

## MARKETING ET INTELLIGENCE ARTIFICIELLE : PIÈGES ET POSSIBILITÉS

**Arnaud de Bruyn**  
Département de Marketing



18

## LE DÉVELOPPEMENT DURABLE GRÂCE AU DATA FOOTPRINT

**Jeroen Rombouts**  
Département des Systèmes d'Information, Sciences de la Décision et Statistiques



20

## LA PRISE DE DÉCISION À L'ÉPOQUE DE L'IA

**Ivana Ljubic**  
Département des Systèmes d'Information, Sciences de la Décision et Statistiques



24

## QUI DEVRAIT POSSÉDER VOS DONNÉES ? UNE ANALYSE ÉCONOMIQUE

**Wilfried Sand-Zantman**  
Département d'Économie



28

## À L'ÈRE DE L'IA, AVONS-NOUS ENCORE BESOIN DE CHERCHEURS ?

**Vivianna Fang He**  
Département de Management



32

## IA ET GOUVERNANCE DE L'INFORMATION : UNE RENCONTRE PROMETTEUSE ?

**Isabelle Comyn-Wattiau**  
Département des Systèmes d'Information, Sciences de la Décision et Statistiques



36

## L'IA AU SERVICE DE L'OCTROI DES CRÉDITS AUX PARTICULIERS : PROMESSES ET DÉFIS

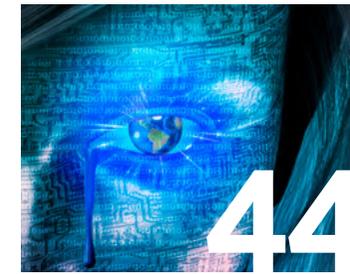
**Andras Fulop et Laurence Daures-Lescourret**  
Département de Finance



40

## COMMENT LES DONNÉES ET LES TALENTS DU DATA SONT ESSENTIELS POUR LE SUCCÈS DU DÉVELOPPEMENT DES ÉNERGIES RENOUVELABLES

**Jeroen Rombouts**  
Département des Systèmes d'Information, Sciences de la Décision et Statistiques



44

## L'INTELLIGENCE ARTIFICIELLE ET ÉTHIQUE

**Laurent Bibard**  
Département de Management



48

## ÉGALITÉ DES SEXES : L'IA EST-ELLE UNE BÉNÉDICTION OU UNE MALÉDICTION ?

**Estefania Santacreu-Vasut**  
Département d'Économie



52

## LES DILEMMES ÉTHIQUES DES SYSTÈMES DE GOUVERNANCE ALIMENTÉS PAR L'IA

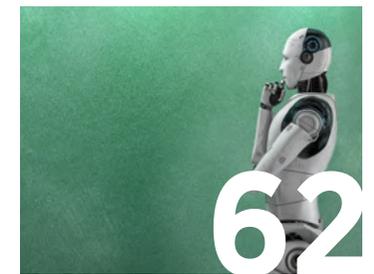
**Julien Malaurent et Thomas Huber**  
Département des Systèmes d'Information, Sciences de la Décision et Statistiques



56

## VERS UNE POLITIQUE DE SÉCURITÉ DES ALGORITHMES ?

**Guillaume Chevillon**  
Département des Systèmes d'Information, Sciences de la Décision et Statistiques



62

## TELLEMENT DE MANIÈRES DE SE TROMPER SUR LE FUTUR DE L'IA, MAIS ELLE NOUS DÉPASSERA

**Fabrice Cavarretta**  
Département de Management

# COMMENT L'IA RÉVOLUTIONNE L'ANALYTIQUE ET CE QUE CELA SIGNIFIE POUR LES ENTREPRISES



**Nicolas Prat** est Professeur associé au Département Systèmes d'Information, Sciences de la Décision et Statistiques (IDS). Il est titulaire d'un doctorat en systèmes d'information et de l'habilitation à diriger des recherches de l'Université Paris-Dauphine, et diplômé de l'ESSEC. Son domaine d'expertise couvre des sujets tels que la business intelligence et l'analytique, les sciences de la conception, la traçabilité et la provenance des données. Ses recherches ont notamment été publiées dans le *Journal of Management Information Systems*, *Decision Support Systems*, et *Expert Systems with Applications*. Il est membre du comité éditorial du *Journal of Database Management*. Il a été chercheur invité à l'Université d'Alicante (Espagne), ainsi qu'à la MIT Sloan School of Management.

Les progrès majeurs en matière d'intelligence artificielle (IA) transforment une grande partie des secteurs d'activité et des métiers. Dans son article intitulé *Augmented Analytics*, le Professeur Nicolas Prat s'intéresse à l'impact qu'a l'intelligence artificielle sur l'analytique. Comment celle-ci révolutionne-t-elle l'analytique et quels sont les défis et opportunités pour les managers ?

L'analytique (analytics) désigne les technologies et processus employés pour collecter, combiner, modéliser, analyser et visualiser les données, afin d'obtenir des informations servant à appuyer une décision. Différents types d'analyse ont pu être définis, notamment : l'analyse descriptive, diagnostique, prédictive et prescriptive.

Le terme "analytics" a gagné en popularité à partir du milieu des années 2000. Au début des années 2010, la révolution Big Data lui a permis de se développer. Les logiciels d'analytique en libre-service donnent la possibilité aux managers de traiter leurs propres données et de créer des visualisations sans devoir systématiquement recourir au département IT. Aujourd'hui, l'IA (et notamment le machine learning et le

traitement automatique du langage naturel) constitue une nouvelle révolution pour ces outils. L'entreprise Gartner parle d'analytique augmentée (augmented analytics). D'autres évoquent la « génération cognitive de l'aide à décision », l'analytique intelligente (smart analytics) ou simplement l'analytique augmentée par l'IA. L'analyse prédictive (et plus généralement les méthodes d'analytique avancées) s'est traditionnellement appuyée sur des algorithmes de machine learning, comme les réseaux neuronaux, pour créer des modèles. En revanche, ce qui est nouveau est que les applications de l'IA interviennent dorénavant à tous les niveaux du cycle analytique. Par exemple, on peut appliquer le machine learning à la recherche du meilleur modèle de machine learning (utiliser le machine learning pour automatiser le machine learning...).

## L'intelligence artificielle imprègne l'ensemble du cycle analytique

Le cycle analytique comprend les phases suivantes :

- Phase 1 : l'identification du problème à résoudre avec l'analytique, ainsi que des opportunités pour l'entreprise.



- Phase 2 : la préparation des données, dans laquelle on distingue le profilage des données (vérification de la qualité) et leur transformation.
- Phase 3 : l'analyse en elle-même. Dans cette phase, on opère une distinction entre la découverte de données (qui peut être du fait des analystes ou des managers utilisant les logiciels de business intelligence en libre-service) et la modélisation (construction et évaluation de modèles par des data scientists).
- Phase 4 : le déploiement des modèles dans les systèmes de production.
- Phase 5 : la prise de décision fondée sur le résultat des analyses.
- Phase 6 : l'action (mise en œuvre des décisions).
- Phase 7 : le contrôle et le suivi des résultats des actions prises.

Les exemples ci-après illustrent quelques cas d'application de l'IA au cours du cycle analytique. Bien qu'elles ne soient pas toutes au même stade de maturité, ces applications sont actuellement mises en œuvre, et d'autres vont encore émerger, à mesure que l'IA continue de se rapprocher de l'intelligence humaine.

Dans la phase 2, certains outils guident les utilisateurs dans la préparation

des données, en suggérant des transformations pour nettoyer ou standardiser les données. Ici, les gains de performances procurés par l'IA peuvent être significatifs puisque, très souvent, cette préparation compte pour plus des deux tiers du temps consacré à l'ensemble du cycle analytique.

Dans la phase 3, le machine learning et le traitement automatique du langage naturel transforment significativement la découverte des données. Ils permettent par exemple un renforcement des visualisations avec des analyses avancées (clustering et prédiction), des conseils en matière de découverte des données (suggestions de visualisations fondées sur l'avancée du processus de découverte de données), l'interrogation de données en langage naturel, ou encore la génération de langage naturel (génération automatique de texte résumant les informations importantes dans une visualisation). Concernant la modélisation, certains outils utilisent le machine learning pour automatiser la génération et l'évaluation des modèles de machine learning, ce qui est traditionnellement un rôle très important des data scientists.

Dans les phases 5 et 6, certaines décisions opérationnelles sont déjà appliquées et mises en place grâce à l'IA, l'exemple le plus caractéristique étant le trading haute-fréquence.

## La démocratisation de l'analytique

Les systèmes interactifs d'aide à la décision (SIAD) sont apparus dans les années 1970. Ils ont progressivement évolué, donnant de plus en plus de liberté aux managers dans l'analyse de leurs données pour obtenir des informations et prendre les bonnes décisions. Les systèmes d'information pour dirigeants (Executive Information Systems, EIS) sont la toute première version des tableaux de bord électroniques. Grâce au développement des entrepôts de données, les managers ont pu créer leurs analyses de façon indépendante, c'est-à-dire sans avoir à solliciter systématiquement le département IT. Avec les logiciels de business intelligence en libre-service, la démocratisation de l'analytique s'est affirmée, à l'aide d'outils comme Tableau, Qlik ou Power BI, facilitant la découverte de données et la visualisation. En revanche, ces outils se sont d'abord cantonnés



aux tâches traditionnelles, à savoir le reporting, le traitement analytique en ligne (OnLine Analytical Processing, OLAP) et les tableaux de bord. La découverte intelligente de données (smart data discovery, un terme introduit par Gartner) a apporté aux outils de business intelligence en self-service des techniques d'analytique avancées. Par exemple, l'outil Watson Analytics d'IBM repose en grande partie sur l'analyse prédictive. Cette découverte intelligente de données a abaissé la barrière à l'entrée pour accéder à l'analytique avancée. Ainsi, spécialistes de domaine, analystes et managers sont devenus, selon l'expression du cabinet Gartner, des citoyens data scientists (citizen data scientists). Avec la dernière génération d'analytique augmentée par l'IA, la démocratisation va plus loin que la simple découverte de données. Ainsi, du moins en principe, vous n'avez plus besoin d'être un professionnel de l'IT pour nettoyer les données, ou un data scientist pour créer et évaluer des modèles de machine learning.

L'affirmation selon laquelle des décisions fondées sur les données entraînent de meilleurs résultats et la maîtrise de l'analytique conduit à de meilleures performances est maintenant communément acceptée et a été étayée empiriquement. À l'ère du Big Data et de l'Internet des objets (Internet of Things, IoT), les données pouvant appuyer une décision sont de plus en plus nombreuses et variées, et les décisions doivent être prises de plus en plus rapidement. Ce contexte, couplé à la pénurie de data scientists sur le marché du travail, demande aux managers de jouer un rôle de plus en plus important dans l'analytique. L'IA peut les y aider. En revanche, la croissance de l'analytique augmentée par l'IA apporte de nouveaux défis, notamment en matière de gouvernance et de qualité des données.

## Qualité des données et gouvernance

La qualité de l'analytique dépend en grande partie de la qualité des données en amont. Dans les faits, les données sont souvent incomplètes, inexactes, contradictoires ou biaisées. Le fait que l'IA permette de démocratiser la préparation des données multiplie les risques de mauvaise qualité des données. De plus, la qualité des données a une influence sur la confiance accordée à l'analytique augmentée par l'IA, de la même manière qu'elle a une influence sur la confiance accordée à l'IA en général. L'intelligence humaine a encore un rôle important à jouer dans la garantie de la qualité des données, notamment dans l'interprétation de ce qu'est une «mauvaise donnée».

Au-delà de la qualité des données, l'analytique augmentée par l'IA met le sujet de la gouvernance de l'ensemble du cycle analytique au centre des préoccupations. Ce cycle est complexe et il peut être composé de plusieurs sous-cycles et instancié de plusieurs façons. Le processus fait intervenir de multiples parties prenantes et outils, et l'accès démocratisé à l'analytique redéfinit les rôles de ces parties prenantes, rajoutant une couche de complexité dans l'orchestration des différentes responsabilités dans le cycle.

Avec la révolution numérique, certaines tâches traditionnellement assignées au département IT se sont déplacées vers d'autres départements. En revanche, ce département garde un rôle essentiel dans la gouvernance de l'IT en général, et dans la gouvernance de l'analytique en particulier, notamment par la définition des rôles, des processus et des normes, établie en lien avec les managers.

## Là où l'intelligence artificielle ne peut pas (encore) battre les humains

En permettant l'automatisation de tâches autrefois réalisées par des agents humains, l'IA entraîne une redéfinition des rôles. Les rôles traditionnels dans le cycle analytique doivent évoluer en prenant en compte les compétences qui permettent aux humains de se distinguer des machines. Bien que les capacités de l'IA progressent à grande vitesse, les humains surpassent encore les machines dans les domaines suivants : poser les bonnes questions (trouver les problèmes plutôt que les solutionner), créer, interagir avec les autres. Ces compétences encore fondamentalement humaines peuvent suggérer certaines évolutions dans le rôle des data scientists, des analystes et des managers dans le cycle analytique. Les data scientists sont menacés par l'IA dans leur fonction de création et d'évaluation de modèles. Ainsi, ils pourraient solidifier leur rôle dans la phase initiale du cycle, c'est-à-dire identifier les questions opérationnelles que résoudront par la suite les modèles de machine learning. Les data scientists tout comme les analystes devraient par ailleurs aiguïser leurs compétences en storytelling. Comme pour rédiger des romans, raconter une histoire à l'aide de données requiert de la créativité. Finalement, les managers pourraient utiliser leurs compétences relationnelles pour faire de l'analytique un processus plus collectif. Contrairement aux SIAD de groupe d'antan (GDSS), les principaux systèmes d'analytique actuels se concentrent principalement sur les individus. En utilisant la réalité virtuelle ou augmentée (analyse immersive), les managers pourraient analyser les données de façon collective, et débattre sur leur signification, avant de prendre une décision commune.

Bien managée, l'analytique augmentée par l'IA procure l'opportunité d'analyser

plus de données, de façon plus efficace, afin de prendre de meilleures décisions. Au-delà des défis mentionnés ci-dessus, plusieurs questions demeurent ouvertes pour de futurs travaux de recherche et pour les praticiens :

- Tous les managers doivent-ils adopter le rôle de citoyen data scientist, ou faudrait-il désigner une certaine catégorie devant endosser ce rôle ?
- Dans quelle mesure l'IA améliore-t-elle l'utilité et la facilité d'usage des systèmes analytiques ?

- Quels sont les facteurs clés de réussite de l'analytique fondée sur l'IA ?
- Quels sont les principaux facteurs affectant la crédibilité des informations découlant de l'analytique augmentée par l'IA, et quel est le niveau de confiance que l'on peut accorder aux décisions que cette dernière prend ou suggère ? ■

Publié initialement le 18 octobre 2018.

### Lecture complémentaire

Prat, Nicolas (2019). Augmented analytics. *Business & Information Systems Engineering*, 61 (3), 375-380.

# INTELLIGENCE ARTIFICIELLE ET RESSOURCES HUMAINES : ET SI ON JOUAIT À PILE OU FACE ?



**Valery Yakubovich** est Professeur au département de Management à l'ESSEC Business School et Senior Fellow au Wharton Center for Human Resources. Ses résultats de ses recherches ont été publiés dans l'American Sociological Review, Human Relations, Organization Science, California Management Review, Harvard Business Review, ainsi que dans d'autres revues et volumes édités. Ses projets actuels explorent les pratiques organisationnelles et de gestion innovantes dans le contexte de la transformation numérique, et la coévolution des carrières et des réseaux sociaux dans les organisations.

**L**a vitesse à laquelle le cœur du sujet de la transformation digitale s'est déplacé du Big Data au Machine Learning puis enfin à l'Intelligence artificielle est assez stupéfiante. Cependant, le gouffre entre discours et réalité du terrain reste important : 41% des PDG estiment qu'ils ne sont absolument pas prêts à utiliser les nouveaux outils analytiques, et seulement 4% disent qu'ils le sont "dans une large mesure" (IBM). Dans son dernier article rédigé avec Peter Cappeli et Prasanna Tambe (Wharton School), Valery Yakubovich, Professeur de management à l'ESSEC, a identifié quatre défis que rencontrent les équipes RH dans la mise en place d'outils analytiques auxquels il offre des réponses pratiques.

## Tout ce qui brille n'est pas nécessairement de l'or

La sophistication croissante de l'IA en matière d'analyse prédictive la rend très intéressante pour les ressources humaines. Elle pourrait être appliquée à diverses pratiques des RH, comme le recrutement et la sélection, la formation et la fidélisation des employés. Mais comme les ressources humaines ont affaire à des personnes, les questions sont importantes et nuancées et n'ont

pas de réponses toutes faites. Le fait que les ensembles de données sur les ressources humaines tendent à être beaucoup plus petits que ceux d'autres domaines, comme les études de marché, et que les techniques de la science des données sont peu performantes lorsqu'il s'agit de prédire des résultats relativement rares, complique encore les choses. Le licenciement d'une personne en raison de ses mauvaises performances est un exemple de résultat qui se produit relativement rarement dans les entreprises, mais qui a des implications importantes pour les individus et la société.

Les problèmes rencontrés par l'IA dans ses applications RH se répartissent en quatre groupes principaux : la complexité des problèmes RH, les petits ensembles de données, les considérations éthiques et juridiques et les réactions des employés. Nous expliquons comment ces problèmes s'appliquent à chaque étape du cycle de vie de l'IA, de la génération des données à l'apprentissage machine, en passant par la prise de décision, et nous incluons des questions à se poser lors de la conception d'une stratégie d'IA pour la gestion des RH.



© Gettyimages - Olycom/Alamy

## Comment faire pour...

### 1. Générer les données

La collecte des données pour votre algorithme d'IA peut être compliquée. Prenez la question apparemment simple « Qu'est-ce qu'un bon employé ? », une question qui devient moins simple lorsque vous creusez un peu plus. Les exigences du poste peuvent être larges et difficiles à spécifier pour un algorithme. Il y a aussi la question du biais : un algorithme d'IA pourrait être capable d'identifier les relations entre les attributs des employés et les performances professionnelles, mais si, par exemple, une entreprise a historiquement embauché et promu des hommes blancs, l'algorithme pourrait prédire que les hommes blancs seront les plus performants et discriminent par inadvertance les autres candidats, même si ces derniers sont hautement qualifiés. La mesure des performances peut également présenter des difficultés : Qui évalue les performances ? Sur quoi la mesure se base-t-elle ? Nous travaillons dans un écosystème interconnecté, de sorte que les performances sont également influencées par des facteurs tels que nos collègues, les ressources professionnelles et la culture de l'entreprise. Idéalement, un

algorithme devrait inclure de multiples indicateurs de performance, mais il est difficile de créer une variable agrégée pour représenter la performance. Par conséquent, ne cherchez pas de mesures parfaites car elles n'existent pas, mais choisissez plutôt des mesures raisonnables et tenez-vous-en à elles.

Il existe également un certain biais de sélection dans l'évaluation des employés, car souvent seuls ceux qui ont été embauchés sont inclus dans l'ensemble de données. La plupart des entreprises ne tiennent pas de registre de toutes les données qu'elles accumulent. Pour constituer un ensemble de données plus important, il faut regrouper les informations provenant de plusieurs sources et au fil du temps, y compris celles des candidats qui sont éliminés à la présélection.

Avant de lancer un nouveau projet de ressources humaines numériques, déterminez les données nécessaires et disponibles qui peuvent être extraites et transférées dans un format utilisable à un coût raisonnable. Le partage des données entre les unités doit devenir une priorité à court terme ; pour évaluer les performances des employés, vous devez intégrer les données commerciales et

financières de l'entreprise. À long terme, investissez dans la normalisation des données et l'intégration des plateformes dans l'ensemble de votre entreprise.

Avez-vous suffisamment de données pour construire un algorithme ? De petits ensembles de données sont souvent suffisants pour identifier les relations de cause à effet, que les gestionnaires doivent comprendre pour pouvoir agir sur la base des informations obtenues. Par conséquent, moins vous avez de données, plus vous aurez besoin de théorie (tirée de la littérature sur la gestion, des connaissances des experts et de l'expérience des gestionnaires). Les expériences aléatoires ne sont pas à négliger pour tester les hypothèses causales.

Si d'autres entreprises mettent leurs données à disposition pour le machine learning, assurez-vous que votre contexte n'est pas trop distinct afin que l'algorithme construit sur des données provenant d'ailleurs soit efficace pour votre propre organisation. Vous pouvez également utiliser les réseaux sociaux comme source alternative de données : certains employeurs les utilisent pour l'embauche, d'autres pour identifier des problèmes tels que le harcèlement. Les acteurs des





RH doivent également tenir compte des considérations relatives à la vie privée et voir dans quelles conditions les données des employés peuvent être utilisées.

## 2. Utilisation de machine learning

Prenons l'exemple de l'utilisation de machine learning dans le processus d'embauche : nous pourrions examiner quelles caractéristiques des candidats ont été liées à de meilleures performances dans le passé et nous en servir pour éclairer nos décisions d'embauche. L'utilisation d'un algorithme de machine learning peut s'avérer plus efficace que les stratégies classiques, mais elle pose un problème d'autosélection : la capacité du modèle à "continuer à apprendre" et à s'adapter aux nouvelles informations disparaît lorsque le flux de nouvelles embauches est limité par les prévisions de l'algorithme actuel. Pour résoudre ce problème, il pourrait être utile de recycler périodiquement l'algorithme en utilisant des données sur les performances des

candidats qui ne correspondent pas à ses critères. Un autre problème possible est que l'utilisation d'algorithmes dans la sélection pourrait réduire l'éventail des variables d'intérêt, masquant potentiellement les véritables relations. Par exemple, si un responsable du recrutement prend sa décision sur la base des notes universitaires, il pourrait alors avoir du mal à trouver un lien entre les notes et les performances, pour la simple raison que les notes des employés varient peu et que la relation n'est donc pas aussi claire.

L'utilisation d'algorithmes dans les décisions relatives aux ressources humaines peut également poser des problèmes éthiques. Par exemple, si nous considérons la différence entre les populations majoritaires et les populations minoritaires, les algorithmes qui maximisent le succès prédictif pour la population dans son ensemble peuvent être moins précis dans la prédiction du succès pour la population minoritaire. La

génération d'algorithmes distincts pour les deux groupes pourrait conduire à de meilleurs résultats, mais aussi à des conflits avec les normes juridiques de traitement disparate. Ainsi, la mise en œuvre efficace des algorithmes d'apprentissage automatique nécessite une révision du droit du travail.

## 3. Prise de décision

Lorsqu'il doit choisir entre deux candidats qui sont tous deux qualifiés pour le poste, le responsable du recrutement doit prendre une décision difficile. Supposons qu'un algorithme détermine que l'un des candidats correspond à 80 % au poste et l'autre à 90 %. Une différence de 10 % est-elle grande ou petite, compte tenu de certaines erreurs de mesure et de certains biais très probables ? Afin d'atténuer certaines de ces questions, nous pourrions introduire la variation aléatoire, qui est un mécanisme non reconnu mais important dans la gestion. Contrairement à la croyance populaire, les recherches montrent que

les employés perçoivent les processus aléatoires comme justes dans la détermination de résultats complexes et donc incertains. Par conséquent, si les deux candidats sont forts, il est plus logique de faire un choix aléatoire. En d'autres termes, la randomisation devrait être un outil de gestion de l'IA.

L'accord des salariés est également un élément clé de l'équation, car ils seront touchés par les changements dans le processus décisionnel. Comment les employés réagiront-ils aux décisions prises par un algorithme plutôt que par un superviseur ? Même si les salariés ne sont pas toujours engagés envers l'organisation, ils peuvent l'être envers leur supérieur. Examinons l'exemple suivant. Sur le lieu de travail, si votre superviseur vous assigne un travail le week-end, vous pouvez le faire sans vous plaindre si vous pensez que votre superviseur est généralement juste. Lorsque l'horaire de travail est généré par un programme, vous pourriez réagir

différemment, car vous n'avez pas de relation préexistante avec l'algorithme. Cela étant dit, certaines décisions sont plus faciles à accepter de la part d'un algorithme, surtout lorsque ces décisions ont des conséquences négatives pour nous, comme une augmentation des prix, car la décision semble moins personnelle.

## Et maintenant, que faisons-nous ?

Pour conclure, voici quelques questions qu'il faut vous poser avant d'utiliser l'IA dans vos processus RH. Souvenez-vous que :

1. Les explications causales sont essentielles pour l'analyse et la prise de décision dans le domaine des RH car elles peuvent garantir l'équité, être comprises par les parties prenantes et sont défendables sur le plan éthique et moral.
2. Les entreprises doivent accepter le pouvoir prédictif relativement faible des algorithmes RH.

3. La randomisation peut aider à établir la causalité et compenser partiellement le faible pouvoir prédictif des algorithmes.

4. La formalisation des processus de développement des algorithmes et l'implication des parties prenantes dans le processus aideront les employés à former un consensus sur l'utilisation des algorithmes et à accepter leurs résultats. ■

*Publié initialement le 29 octobre 2018 ; mis à jour en décembre 2020.*

### Pour aller plus loin

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, 61 (4), 15-42.



# MARKETING ET INTELLIGENCE ARTIFICIELLE: PIÈGES ET POSSIBILITÉS



**Arnaud De Bruyn** est Professeur au département de Marketing. Les recherches d'Arnaud portent sur le marketing analytique, le marketing interactif et direct, ainsi que sur l'ingénierie marketing. Il a publié notamment dans Marketing Science, Management Science, et Information Systems Research. Arnaud a reçu plusieurs prix scientifiques, notamment pour ses recherches sur le marketing viral. Il est également Distinguished Visiting Research Scholar à l'Institute for the Study of Business Markets. Arnaud a également une grande expérience du secteur non lucratif, où il a conseillé plus de quarante organisations à but non lucratif pour améliorer leur marketing direct. Arnaud a un doctorat en sciences du management de la Penn State University.

Article écrit par Julia Smith, rédactrice en chef, ESSEC Knowledge

L'ESSOR de l'intelligence artificielle a naturellement vu les gens l'appliquer (ou tenter de l'appliquer) d'innombrables façons, avec plus ou moins de succès. Si l'intelligence artificielle peut être un outil puissant - entre de bonnes mains, dans la bonne situation -, elle n'est pas aussi facile à mettre en œuvre qu'un système d'information avec un simple pression sur quelques touches. Cela est peut-être particulièrement vrai pour les environnements qui traitent du comportement humain, comme le marketing. Comme le dit le proverbe, « à grand pouvoir, grande responsabilité » : et les responsables marketing doivent être conscients de ses pièges potentiels pour éviter les problèmes. Tout aussi important est le besoin de savoir comment déployer correctement leurs outils d'IA pour éviter de gaspiller à la fois son potentiel et les efforts et ressources de leur entreprise. En comprenant les pièges de l'IA, les responsables marketing peuvent tirer le meilleur parti de ses opportunités.

Jusqu'à présent, les plus grandes avancées de l'IA dans le monde des affaires ont été liées au deep learning, c'est-à-dire aux réseaux neuronaux complexes et multicouches (c'est-à-dire

profonds), à la résolution de problèmes difficiles grâce à l'analyse prédictive. Plus un réseau de neurones comporte de couches, plus il est complexe, et plus les réseaux « stratifiés » peuvent identifier et apprendre des relations plus complexes entre les variables. Cela signifie que l'intelligence artificielle peut apprendre à découvrir des relations que les techniques statistiques existantes ne peuvent pas détecter et qu'elle peut apprendre à le faire de manière autonome. C'est le principal argument de vente des algorithmes d'IA contemporains.

Si la capacité des algorithmes d'IA à créer des modèles de manière autonome est sa force, elle n'est pas sans poser de problèmes lorsqu'il s'agit de la mettre en œuvre. Ces défis sont : un manque de bon sens, des fonctions objectives, un environnement d'apprentissage sûr et réaliste, des algorithmes biaisés, une intelligence artificielle compréhensible et contrôlable, le paradoxe de l'automatisation et le transfert de connaissances.

## Manque de bon sens

Qu'entendons-nous par manque de bon sens ? Ce n'est pas une insulte à



ses programmeurs ou aux utilisateurs ; non, nous voulons dire que l'algorithme lui-même manque de ce que nous, les humains, appelons le « bon sens ». Nous savons que l'intelligence émotionnelle est importante, et en effet les systèmes d'IA sont de plus en plus capables de reconnaître les émotions des gens, par la reconnaissance d'images, l'analyse de la voix ou l'analyse de textes. Mais reconnaître les émotions ce n'est pas les comprendre et les ressentir. Un système d'IA pourrait apprendre que les mots "reine" et "couronne" sont liés, et pourrait même les utiliser de manière appropriée dans une phrase, mais le sens des mots et des phrases s'en trouverait perdu. Tout ce qui s'approche du bon sens doit être programmé par une personne, ce qui devient un problème lorsqu'il s'agit de fonctions objectives.

## Fonctions objectives

Une fonction objective est une fonction qui spécifie le résultat que l'algorithme d'IA vise à optimiser (Sutton et Barto, 2018). Dans le contexte du marketing, cela pourrait ressembler à une maximisation des profits ou à la fidélisation des clients. La « liberté » de l'IA par rapport au bon sens entrave sa capacité à définir une

fonction objective. Il se peut que les humains comprennent quelque chose de manière implicite, mais qu'ils aient ensuite du mal à traduire cela pour l'algorithme. Cela pourrait mal tourner : une voiture autonome qui a pour mission de « se rendre à l'aéroport dès que possible » pourrait y arriver en un temps record, mais en ayant fauché des piétons et grillé des feux rouges sur son chemin. Si l'exemple précédent est évidemment extrême, nous avons déjà vu les conséquences de ce jeu dans la vie réelle, avec des systèmes fondés sur le sexe ou la race. Un résultat comme la maximisation des profits ne peut être envisagé sans tenir compte des implications juridiques, morales et éthiques, que les acteurs du marketing doivent garder à l'esprit lorsqu'ils élaborent et mettent en œuvre leurs systèmes.

## Un environnement d'apprentissage sûr et réaliste

Comme vous pouvez l'imaginer, tout cela est plus facile à dire qu'à faire. Le transfert de connaissances de l'expert à l'algorithme et vice versa est l'un des plus grands problèmes auxquels l'IA est confrontée aujourd'hui, et le potentiel

d'erreurs coûteuses est énorme. Pour éviter les retombées, il est important que les algorithmes d'IA apprennent dans un environnement sûr et réaliste. Sûr, car s'ils font des erreurs, l'impact sur l'entreprise est moindre et ils évitent l'équivalent marketing d'un feu rouge. Réaliste, car les données ressemblent à ce qu'ils recevraient dans une situation réelle. Cela représente un défi pour le marketing, car les clients peuvent être imprévisibles, et un nouveau facteur (comme, par exemple, la COVID-19) peut mettre à mal les campagnes de marketing les mieux conçues. S'il peut être tentant de penser que l'IA réduit, voire élimine, notre besoin de comprendre le comportement des clients, c'est le contraire : nous avons plus que jamais besoin d'une théorie détaillée du comportement des clients, car cela nous aidera à mieux configurer nos algorithmes d'IA.

## Algorithmes biaisés

Cela nous amène à une autre limitation de l'utilisation de l'IA dans le marketing : son potentiel à être biaisé. Bien sûr, l'algorithme lui-même n'est pas biaisé, mais s'il est suffisamment puissant, il pourrait identifier une caractéristique comme la race ou le sexe par lui-même et





**Pour aller plus loin**

De Bruyn, A., Viswanathan, V., Shan Beh, Y., Kai-Uwe Brock, J., & von Wangenheim, F. (2020). Artificial intelligence and marketing: Pitfalls and opportunities. *Journal of Interactive Marketing*, 51, 91-105.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction Second Edition*. Cambridge, MA: MIT Press.

faire des prédictions biaisées. Comment cela se fait-il ? Il peut capter d'autres informations qui agissent comme un proxy du facteur en question, comme l'éducation ou le revenu, reproduisant ainsi involontairement les biais que l'on trouve dans les données. Dans le contexte du marketing, cela pourrait conduire à des résultats tels qu'un algorithme d'optimisation des prix visant à faire payer davantage les femmes ou un algorithme de publicité ciblant une population vulnérable. Cela a des implications légales aussi bien qu'éthiques évidentes. Le problème est compliqué par le fait que l'ajout de la variable sociodémographique en question au modèle pour tenter de le clarifier pourrait simplement faciliter la

réalisation de prédictions biaisées par l'algorithme. Si les acteurs du marketing ne comprennent pas correctement les algorithmes qu'ils utilisent, ils pourraient ne pas savoir comment contester ces prédictions troublantes.

**L'intelligence artificielle compréhensible**

La capacité à comprendre et à expliquer le modèle est un autre facteur de l'adoption de l'IA. Si vous utilisez un modèle d'IA, vous devez comprendre pourquoi il fait les prévisions qu'il fait et être capable d'interpréter ce que fait le modèle. Plus précisément, les « manipulateurs » humains de l'IA doivent être capables

d'expliquer : 1) l'objectif du modèle, 2) les données qu'il utilise et 3) le lien entre les entrées et les sorties. En comprenant cela, il est également possible de savoir pourquoi le système d'IA est préférable à un système non-IA.

**Intelligence artificielle contrôlable**

L'utilisation du terme "handlers" ci-dessus était intentionnelle : un système d'IA doit pouvoir être contrôlé et neutralisé. Cela pourrait faire apparaître des images de moi, de I, Robot et de robot tueur, et bien que la réalité soit moins mortelle, elle reste sérieuse. Un exemple récent est que l'algorithme

de tarification d'Uber a répondu à l'afflux de personnes fuyant les lieux de l'attentat terroriste de juin 2017 à Londres en adaptant (en augmentant donc) le prix des trajets à plus du double du tarif habituel. Tous ceux qui ont pris Uber connaissent malheureusement leur système de tarification de pointe, mais à la suite d'un attentat terroriste, Uber est passé pour un profiteuse impitoyable. Cependant, le système de surveillance d'Uber a rapidement signalé le problème, et ils ont mis en place des mécanismes qui leur ont permis de passer outre l'algorithme en quelques minutes. Ils ont également été rapides à communiquer sur ce qui se passait, ont fait des tours gratuits dans cette zone et ont remboursé les personnes

touchées. Hélas, le mal était fait. Cette situation a laissé une marque noire sur leur réputation et sert d'avertissement aux responsables marketing que tout algorithme qu'ils mettent en œuvre doit être constamment surveillé et avoir la possibilité d'être contourné.

**Le paradoxe de l'automatisation**

L'objectif de l'automatisation est de remplacer le rôle de l'homme, en visant à rendre les tâches plus rapides et plus précises et en laissant les gens libres d'effectuer des travaux plus complexes. L'inconvénient est que les gens n'ont pas l'expérience de ces tâches plus simples et n'ont pas la possibilité de développer progressivement leur expertise et leurs compétences. Dans le domaine du marketing, cela pourrait signifier que les personnes travaillant dans ce domaine, des agents du service clientèle aux analystes des études de marché, ratent l'occasion d'affiner leurs compétences sur des tâches plus simples et plus répétitives qui leur permettent de mieux comprendre les clients et leurs besoins, et se retrouvent à ne traiter que les cas les plus compliqués et les plus uniques. Il reste à voir quelles seront les conséquences de cette situation sur la qualité du service et du travail.

**La prochaine frontière de l'IA et du marketing : transférer et créer des connaissances**

Ce qui distingue l'IA des statistiques traditionnelles est sa capacité à effectuer un apprentissage de haut niveau, comme la mise en évidence de relations entre des indicateurs pour prédire la probabilité qu'un internaute clique sur une publicité, et ce de manière autonome. Pouvoir créer des connaissances de ce type est un énorme avantage de l'IA. Cependant, le transfert de connaissances du modèle d'IA à

l'expert et vice versa est une faiblesse majeure de l'IA. Comme le marketing traite du comportement humain, cela nécessite beaucoup de bon sens, ce qui, comme nous le savons maintenant, n'est pas le point fort des modèles d'IA. Comme ce type de connaissance est souvent plus implicite, traitant des codes et des normes sociales, il est également plus difficile de programmer dans un modèle d'IA. La machine sera également capable de repérer les liens qu'elle doit transférer à l'expert humain, notamment pour que les experts puissent identifier les failles du système et comprendre son fonctionnement. Un système d'IA capable de créer et de transférer des connaissances à l'expert humain est donc le Saint Graal de la technologie de l'IA. ■

**Les points à retenir**

Que peut donc faire un responsable marketing qui souhaite utiliser l'IA ? Il y a quelques points clés à garder à l'esprit :

1. Comprendre l'objectif de la mise en œuvre du système d'IA. Quel est votre objectif ?
2. Identifier la valeur ajoutée du système d'IA. Qu'apporte-t-il en plus des capacités humaines ?
3. Comprendre ce que fait votre système d'IA. Quelles sont les données qu'il analyse ? Comment produit-il les résultats ?
4. Examinez le système pour voir s'il n'est pas biaisé. Votre système comporte-t-il des biais intégrés ?
5. Communiquer : veiller à ce que les parties prenantes concernées (consommateurs, employés) aient la possibilité d'observer le système d'IA et d'interagir avec lui, d'établir la confiance, d'assurer le transfert réciproque de connaissances et de pratiques.

Publié initialement le 4 janvier 2021



# LE DÉVELOPPEMENT DURABLE GRÂCE AU DATA FOOTPRINT



**Jeroen Rombouts** est Professeur au département de Statistiques et d'économétrie à l'ESSEC Business School depuis 2013. Il est le titulaire de la chaire Accenture Strategic Business Analytics. Il a obtenu son doctorat en économétrie en 2004. Il donne des cours d'analyse de données & de science des données aux programmes MBA et masters à Paris & à Singapour. Il étudie l'analyse prédictive, en particulier l'analyse des séries chronologiques et les prévisions. Il a publié de nombreux articles scientifiques, organise des ateliers et est rédacteur en chef adjoint de plusieurs revues scientifiques quantitatives. Ses recherches sont financées par la Commission européenne et l'Agence nationale de la recherche. Il intervient fréquemment en tant que consultant expert sur des sujets d'analyse stratégique d'entreprise et de science des données. Avant de rejoindre l'ESSEC, il était professeur associé à HEC Montréal.

## Co-auteur

**Reda Gomery:** Reda Gomery est vice-président de data à Kearney. Contact : Reda.Gomery@kearney.com

Dès le début de la pandémie COVID-19, les entreprises ont dû accélérer leur transformation numérique. Cela implique des investissements accrus, si importants qu'ils nécessitent un soutien des décideurs de l'organisation. Les enjeux sont majeurs pour les organisations. De l'accélération des ventes à l'optimisation des processus opérationnels, le numérique a un impact sur la chaîne de valeur dans tous ses aspects. Si la révolution numérique génère une modernisation inévitable des entreprises et un espoir de création de valeur, elle provoque également un défi majeur pour les organisations : les données.

Les données des transactions, des clients, des produits, etc. envahissent le fonctionnement quotidien des organisations, constituant un atout potentiellement précieux, mais surtout un défi important en termes de gouvernance et de management. Les organisations doivent améliorer la compréhension de ces données dans le cadre de leur transformation.

À très court terme et en ces temps peu certains, les données deviennent plus que jamais indispensables pour identifier les leviers de performance des entreprises. L'optimisation des coûts, l'augmentation des revenus des entreprises et l'efficacité

des processus sont des initiatives fondées sur la disponibilité de données pertinentes. Avec l'accélération des cycles de décision, de nombreux décideurs ne seront plus en mesure de piloter leurs entreprises avec des données approximatives et souvent inexactes. Disposer de bonnes données - et juste à temps - est devenu une nécessité urgente. Mais cette perspective ne semble réalisable que si le patrimoine de données est mieux maîtrisé. C'est précisément l'objectif de la méthode « Data Footprint » conçue par Kearney et l'ESSEC. L'évaluation de l'empreinte des données constitue désormais une approche essentielle pour sécuriser les investissements et accroître le contrôle du patrimoine de données.

L'approche Data Footprint introduit une pratique vertueuse qui vise à comprendre le patrimoine de données, les risques, les défis et les limites liés aux données au sein des organisations. Le Data Footprint est un processus d'évaluation basé sur une analyse à 360° des données requises dans le cadre d'une initiative d'entreprise pilotée par l'entité en charge de la gouvernance des données. L'objectif du Data Footprint est d'évaluer les actifs de données afin d'établir un score d'évaluation des risques. Basée sur de multiples dimensions d'analyse telles que la qualité ou la sécurité des données,



© Gettyimages - Fotopostok

notre méthode permet une évaluation quantifiée du patrimoine de données dans une organisation. Aujourd'hui, le patrimoine de données est encore mal contrôlé et exploité dans de nombreuses entreprises. Quel est le niveau de qualité des ensembles de données critiques dans l'organisation (par exemple, les données des clients/fournisseurs...)? Quel est le niveau de risque associé? Quel est le degré de contrôle et de propriété de ces données dans l'organisation? Ces questions sont souvent posées par les décideurs sans réponses concrètes basées sur une évaluation structurée. La complexité des systèmes d'information combinée à l'absence de gouvernance rend l'équation des données souvent complexe et coûteuse.

Le Data Footprint permet aux entreprises d'obtenir une évaluation tangible des données dans plusieurs dimensions afin d'établir un score de risque. L'objectif d'une telle mesure est de pouvoir évaluer avec précision les points faibles et de suivre les améliorations du patrimoine de données. L'approche permet également d'établir des points de référence internes et externes sur la base d'une grille d'analyse standardisée.

La mise en œuvre du Data Footprint doit être progressive tout en se concentrant sur les ensembles de données critiques dans le

contexte des grands programmes, projets ou initiatives de transformation des entreprises.

La démarche doit impliquer plusieurs collaborateurs, au moins des représentants des métiers et de l'informatique, qui utilisent conjointement une fiche de score basée sur les cinq dimensions suivantes : accessibilité et disponibilité ; qualité ; propriété ; risques ; et identification des futurs utilisateurs. Le score global calculé sur ces cinq dimensions peut varier entre 0 et 15, plus le score est faible, plus le risque lié à l'initiative de l'entreprise est élevé.

Prenons l'exemple d'une entreprise spécialisée dans la distribution de matériel électronique au grand public grâce à son réseau de distribution de plus de 2000 magasins. Dans le cadre de sa stratégie en matière de données, l'entreprise décide de lancer un projet prioritaire qui déploie une approche centrée sur le client afin d'accroître la valeur du client. L'objectif est de mieux comprendre les préférences des clients afin de répondre à leurs attentes. L'entreprise anticipe un risque potentiel important lié aux données (disponibilité, qualité, etc.) et décide de lancer une approche de « Data Footprint ».

Le score total de risque de Data pour cette entreprise était inférieur à 5 lors de l'exercice

d'évaluation. Sur la recommandation du Chief Data Officer et en accord avec le reste de l'équipe, la décision de lancer le projet est reportée en attendant la mise en œuvre d'un plan d'action spécifique lié aux données. Cette approche a permis à l'entreprise d'appréhender un risque majeur lié aux données de ce projet. En effet, un lancement rapide de ce projet sans évaluation préalable aurait potentiellement conduit à un échec avec des conséquences économiques (pertes estimées à quelques centaines de milliers d'euros). La démarche a également permis d'initier un travail collaboratif autour des données sur toute la durée de cette évaluation (un mois), et ainsi d'éviter des malentendus internes sur les responsabilités des différents intervenants (Business lines, équipes informatiques, etc.). Enfin, un plan d'action clair a pu être établi, justifiant l'investissement de ressources techniques et humaines pour faire évoluer le système d'information. ■

**Pour une version plus technique de cet article ou pour plus de détails sur le Data Footprint, veuillez contacter :**

Reda Gomery, vice-président, Kearney, Reda.Gomery@kearney.com  
Jeroen Rombouts, professeur, Essec Business School, rombouts@essec.edu

Publié initialement le 19 octobre 2020



# LA PRISE DE DÉCISION À L'ÉPOQUE DE L'IA



**Ivana Ljubic** est Professeure à l'ESSEC Business School où elle est la directrice académique du programme EMBA de l'ESSEC et de Mannheim. Elle enseigne l'aide à la décision, la recherche opérationnelle et la gestion de la chaîne d'approvisionnement. Elle a été professeur ou chercheur invité à l'Université de Vienne, à la Robert H. Smith School of Business, à l'Université technique de Dortmund, à l'Université technique de Berlin et à l'Université Dauphine. Elle enseigne et effectue des recherches sur le rôle de la recherche opérationnelle dans la prise de décision à l'ère des grandes données et de l'IA. Elle a obtenu un doctorat en informatique à l'Université technique de Vienne et une habilitation en recherche opérationnelle à l'Université de Vienne. Elle a publié plus de 50 articles dans des revues scientifiques dans le domaine de la recherche opérationnelle & de management.

Le terme « IA » existe depuis près de 70 ans, mais ce n'est qu'aujourd'hui que l'IA affecte et change profondément notre façon de vivre, de travailler, d'interagir et/ou de jouer. L'IA est en train de révolutionner et de perturber diverses industries, et les dirigeants d'entreprises adoptent cette tendance avec beaucoup d'enthousiasme. Alors, pourquoi n'avons-nous pas vu l'essor de l'IA beaucoup plus tôt et est-elle là pour rester cette fois-ci ? Ou est-ce que l'IA est encore une nouvelle technologie qui surfe sur la vague de la mode ? Et comment peut-elle nous aider à mieux prendre les décisions ?

## Tout d'abord, qu'est-ce que l'IA ?

Une grande partie de chercheurs sont d'accord avec Larry Tesler (un informaticien qui a inventé le copier-coller alors qu'il était chez Xerox et a travaillé sur l'interaction personne-machine) pour dire que « l'IA est tout ce qui n'a pas encore été fait ». Cela signifie que la définition de l'IA change au fil du temps. Au fur et à mesure que nous nous habituerons aux progrès antérieurs de la technologie et de l'informatique, ce

que l'on appelle aujourd'hui l'IA sera considéré comme normal demain.

Néanmoins, beaucoup conviendront que, contrairement aux nombreux « hivers de l'IA » que nous avons connus dans le passé, la situation est différente aujourd'hui : le temps est venu et elle est là pour rester. Cela est dû aux trois principales forces qui ont convergé pour permettre la « révolution de l'IA » à laquelle nous sommes confrontés aujourd'hui :

- 1) L'augmentation des capacités de traitement et de stockage : la puissance de calcul a été multipliée par mille milliards entre 1956 et 2015, avec la baisse rapide du coût des technologies, une tendance qui se poursuit à ce jour. Les consoles Nintendo de 1983 sont aussi puissantes que l'ordinateur de guidage Apollo, qui a amené les premiers humains sur la lune en 1969. Et l'iPhone 4 en 2010 est aussi puissant que l'ordinateur le plus rapide du monde en 1985, le supercalculateur Cray-2.
- 2) Cloud computing : le stockage des données et les capacités informatiques sont désormais externalisés vers les « nuages ». Aujourd'hui, des concepts



tels qu'Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) et même Machine-Learning-as-a-Service (MLaaS) sont disponibles à des prix abordables. Ils offrent de faibles barrières à l'entrée et permettent aux entreprises d'adopter rapidement les nouvelles technologies informatiques.

- 3) Connectivité et disponibilité des données en temps réel : de grandes quantités de données variées et en temps réel peuvent être collectées instantanément, grâce aux téléphones mobiles, aux capteurs, à l'Internet des objets et à d'autres dispositifs reliés à l'Internet ou aux réseaux intranet. Cela permet la prévision en temps réel, la prescription de conseils, l'automatisation et la coordination de nombreux processus complexes.

## Comment pouvons-nous améliorer la prise de décision en tirant parti du big data et de l'IA ?

Vous pourriez être enthousiasmés par les récentes avancées technologiques en matière de robotique, des véhicules autonomes, de taxis volants, etc. Vous

pourriez être tout aussi terrifiés par les applications militaires de l'IA, ou par les menaces potentielles liées à la reconnaissance faciale ou au profilage racial. Pour l'instant, concentrons-nous sur une question plus pratique : comment pouvons-nous améliorer les pratiques commerciales quotidiennes en tirant parti de l'analyse des (grandes) données et de l'IA ? C'est là que l'IA combinée à la Business Analytics (IA+BA), également appelée « AI-powered analytics », arrive.

La Business Analytics est l'exploration et l'exploitation des données d'une organisation, avec un accent sur 1) l'analyse des données, 2) les prévisions et 3) l'analyse prescriptive (la recommandation des décisions et l'aide à la décision). Les entreprises qui utilisent la BA s'engagent à prendre des décisions meilleures, plus éclairées et fondées sur des données. Elle est applicable dans de nombreux secteurs différents : commerce de détail, transport, divertissement, soins de santé et/ou énergie, pour n'en citer que quelques-uns. C'est une approche pluridisciplinaire qui combine l'informatique, les statistiques, les mathématiques, la prise de décision et l'optimisation. La BA permet aux organisations d'éradiquer certains biais cognitifs typiques de la

prise de décision. Et, plus important encore, grâce à elle, la prise de décision est basée sur les données et non plus la place dans la hiérarchie.

Dans les organisations, il y a généralement trois niveaux de prise de décision :

- stratégique (impliquant des décisions à long terme, non routinières et complexes prises par la direction générale),
- tactique (concernant les décisions à moyen terme, moins complexes, prises par l'encadrement intermédiaire), et
- opérationnel (consistant des décisions de routine quotidiennes).

L'IA+BA peut être utilisée avec succès à tous ces niveaux, mais avec un niveau d'automatisation différent.

Prenons l'exemple de Netflix : saviez-vous que Netflix suit non seulement les films que vous regardez, mais aussi les résumés que vous lisez et le temps que vous passez à surfer sur les titres et à regarder les bandes-annonces ? En d'autres termes, Netflix connaît le contenu que vous aimez avant que vous le connaissiez vous-même. Il est bien connu que la plateforme utilise des systèmes de recommandation et





des files d'attente personnalisées. Les systèmes de recommandation sont en effet une procédure standard pour la distribution de médias numériques aujourd'hui. Les recommandations fournies aux utilisateurs finaux peuvent être considérées comme des décisions opérationnelles, entièrement automatisées par des algorithmes et des machines, et en tant qu'utilisateur, vous pouvez vous sentir dépassé ou satisfait. Mais il ne s'agit là que de la partie « IA » ou « prédictive » de l'approche IA+BA de Netflix. Qui plus est, Netflix utilise l'IA non seulement pour suggérer ce que vous devriez regarder, mais aussi pour décider quels projets financer. Ainsi, d'importantes décisions stratégiques, comme la budgétisation des investissements, sont prises sur la base de prédictions dérivées de l'énorme quantité de données collectées auprès des utilisateurs. Ainsi, la société utilise une modélisation mathématique et des algorithmes avancés pour décider du contenu à produire (quels genres de films ou types de personnages seront les plus

populaires?), du montant du budget consacré aux différentes productions, et de la programmation des projets (quand et où tourner, avec quelles ressources disponibles, et quand prévoir les dates de sortie?) Quel est le succès de cette approche? Les chiffres parlent d'eux-mêmes: le taux de fidélisation des clients est de 91% (contre 64% pour Hulu et 75% pour Amazon Prime), et le nombre total d'utilisateurs est passé de 20 millions en 2011 à 186 millions en 2020!

Aujourd'hui, de nombreuses décisions opérationnelles et de routine peuvent être entièrement automatisées. La société UPS a utilisé les outils de recherche opérationnelle (la partie « prescriptive » de BA) pour identifier automatiquement les itinéraires les plus efficaces pour les conducteurs de UPS. UPS a intégré ces outils dans les systèmes de navigation des conducteurs et a pleinement déployé la technologie aux États-Unis en 2016. Les résultats ont permis d'économiser presque 40 millions de litres de carburant en un

an, soit l'équivalent de 100 000 tonnes d'émissions de carbone!

L'entreprise parisienne Decision Brain utilise l'IA+BA pour optimiser l'un des plus grands systèmes de location de vélos publics au monde. Le système de location de vélos de Londres utilise près de 12 000 vélos et 800 stations d'accueil et représente plus de 10 millions de trajets par an. Decision Brain utilise des outils d'apprentissage automatique pour la prévision en temps réel des demandes des utilisateurs, et applique des algorithmes d'optimisation pour la gestion des stocks et le déplacement des vélos. Leurs algorithmes font en sorte que lorsque vous arrivez à une station, il y a suffisamment de vélos à louer et suffisamment de places vides pour le retour des vélos.

### Quand est-ce que l'humain doit intervenir ?

Comme le montrent les exemples ci-dessus, il ne fait aucun doute que les

AI-powered analytics aient un potentiel énorme pour tirer parti de l'avantage concurrentiel des entreprises ou les aider à mener des opérations durables qui respectent l'environnement.

Cependant, il y a des décisions importantes pour lesquelles nous ne voulons pas que le jugement humain soit entièrement remplacé par les algorithmes. Un exemple évident est le cas des décisions à long terme qui peuvent impliquer de nombreuses parties prenantes ou nécessiter une analyse minutieuse des aspects géopolitiques ou autres aspects stratégiques. Dans ce cas, l'approche IA+BA basée sur les données peut fournir une analyse des résultats possibles pour plusieurs scénarios futurs (le pire, le prévu et le meilleur, par exemple). Toutefois, ces résultats ne peuvent être considérés que comme des recommandations à l'intention du conseil d'administration, et non comme des décisions à mettre en œuvre immédiatement.

De même, la prise de décision automatique peut entraîner des conséquences involontaires, surtout lorsqu'il s'agit d'élaborer des politiques sensibles qui affectent profondément des vies humaines. Il est très tentant de laisser l'IA analyser la grande quantité de données disponibles pour produire diverses mesures et transformer la vie des gens en scores qui prédisent des résultats éthiquement ambigus comme le potentiel criminel de certains quartiers (police prédictive), la productivité des employés (IA pour les RH) ou la récidive (lors de la prise de décisions de condamnation au tribunal). Cependant, il est important de savoir que ces algorithmes sont très sensibles aux données d'entrée et qu'ils peuvent facilement amplifier la discrimination et être transformés en « armes de destruction mathématique », comme l'explique Cathy O'Neil dans son livre du même titre.

Cependant, il est important de savoir que de nombreux algorithmes sont faits avec

les meilleures intentions et qu'aucun algorithme n'est mauvais en soi! Les algorithmes peuvent coder des préjugés ou des malentendus humains, et dans nos économies accélérées et basées sur les données, cela peut conduire à des inégalités accrues ou à la perpétuation des préjugés raciaux et sexistes, avec des conséquences imprévisibles.

C'est pourquoi les chefs d'entreprise d'aujourd'hui doivent comprendre les défis, les opportunités et les limites de l'IA et de l'analytics. Les nouvelles générations de managers formés par l'ESSEC apprennent le codage — non pas pour devenir des ingénieurs en logiciel ou des spécialistes des données, mais pour comprendre la « pensée algorithmique » et l'importance de développer des outils de décision explicables et éthiques basés sur l'IA. ■

*Publié initialement le 30 septembre 2020*



# QUI DEVRAIT POSSÉDER VOS DONNÉES? UNE ANALYSE ÉCONOMIQUE



**Wilfried Sand-Zantman** a rejoint l'ESSEC Business School en 2020 et était auparavant professeur d'économie à Toulouse School of Economics. Il est titulaire d'un master en économie et statistiques de l'ENSAE (Paris) et d'un doctorat en économie de l'Université de Toulouse I. Les recherches de Wilfried portent sur l'organisation industrielle, l'économie numérique et, plus généralement, l'économie de l'information. Ses articles ont été publiés dans des revues telles que *l'International Journal of Industrial Organization*, *the Journal of Industrial Economics*, *Management Science*, *the Journal of Economics* et *the Journal of European Economic Association*. Wilfried a également été impliqué dans le conseil et la formation des cadres pour plusieurs entreprises telles que EDF, Orange, SFR et Telefonica.

Les données sont devenues un enjeu économique majeur. Dans l'écosystème numérique, la récolte, le traitement et la revente d'information sont les éléments structurants des stratégies des firmes. Que ce soit en fournissant un service gratuit en échange d'une attention ou bien en récoltant des données à l'occasion d'une transaction, de nombreuses firmes utilisent les données fournies de façon plus ou moins consciente et plus ou moins volontaire par les consommateurs. Dans la mesure où ces données sont un actif dans le processus de production, la question du contrôle de ces données et donc du droit de propriété sur ces données se pose.

En Europe (avec le règlement général sur la protection des données) comme aux États-Unis (avec le California Consumer Privacy Act), des réponses juridiques ont été apportées pour permettre aux consommateurs de mieux contrôler la collecte et l'usage de leurs données. Mais la question de la propriété, et donc potentiellement de la possibilité de monétiser d'une façon directe les données personnelles, est encore abordée de façon détournée (voir Duch-Brown et coll., 2017). D'une façon paradoxale, il existe des data brokers qui opèrent sur un marché

des données assez actif (voir FTC, 2014) et rien n'interdit à un individu de vendre ses données privées en échange d'une rémunération. Mais les données ne peuvent pas vraiment être cédées de façon traditionnelle, car le consommateur peut (au moins en Europe) empêcher un tiers d'en faire usage, quand bien même il en aurait autorisé l'usage auparavant.

Malgré ces difficultés pratiques, la réflexion progresse sur ce sujet, bien sûr dans le champ juridique, mais également du côté des économistes. Au-delà de l'étude des conséquences de la mise en place d'un marché des données, il s'agit de comprendre l'impact de l'allocation des droits de propriété soit aux consommateurs, soit aux entreprises qui ont permis d'extraire ces données. Il y a au départ l'idée que les données comme objet économique sont le résultat d'une interaction entre deux parties. Comme elles sont à la fois input et output, les données ont un statut particulier. Comme pour d'autres biens, l'attribution des droits à l'une ou l'autre partie modifie la manière dont les gens consomment ou produisent les biens (voir Coase, 1960).

Dans un article récent (Dosis et Sand-Zantman, 2019), nous proposons



une exploration théorique de cette question dans un cadre de marché biface. Plus précisément, nous analysons une situation dans laquelle des consommateurs consomment un service et les données générées lors de cette transaction pouvant être valorisées par la suite (offre personnalisée, vente à des data brokers, utilisation en interne...). Deux hypothèses importantes sont utilisées dans cette étude. Tout d'abord, la valeur marchande des données liées à un consommateur est d'autant plus grande que son usage du service est intensif et que ces données ont été analysées par l'entreprise. Ensuite, nous supposons que les consommateurs redoutent l'exploitation des données, ce qui a un impact négatif sur leur satisfaction lors que la consommation du service. Dans ce cadre, nous mettons en avant un arbitrage entre deux formes d'inefficacité : la surexploitation des données d'un côté et le sous-investissement de l'autre.

Lorsque les entreprises possèdent les droits des données de leurs consommateurs, comme c'est de facto le cas actuellement dans la plupart des pays, elles ont la possibilité de les valoriser et donc sont incitées à la fois à en générer beaucoup, mais également à la traiter efficacement pour en extraire

la plus grande valeur. Cependant, les consommateurs sont conscients du risque de surexploitation de leurs données et ont tendance à restreindre leur utilisation du service en question, en particulier lorsqu'ils auraient souhaité l'utiliser intensivement en l'absence de ce risque. On a donc un choix efficace de traitement de données, mais un usage limité du service par les consommateurs. Lorsque les consommateurs possèdent les droits sur leurs données, ils peuvent ajuster leur revente éventuelle après avoir considéré l'ensemble des bénéfices et des coûts. Mais l'impossibilité pour les entreprises de récupérer la valeur marchande complète des données les dissuade de les traiter efficacement et donc de créer la valeur marchande la plus élevée pour les consommateurs.

Comment alors faire le choix optimal entre ces deux régimes de propriété ? Même si les intérêts des entreprises et des consommateurs ne sont pas totalement alignés, on peut montrer qu'ils évoluent dans le même sens lorsque le rapport entre la valeur créée sur le premier marché — celui du service, là où la donnée a été générée — et celle créée sur le second marché — celui sur lequel la donnée est revendue — est modifié.

Plus précisément, si la valeur marchande des données sur le second marché est faible, les entreprises gagnent peu à en extraire toute la valeur. Mais l'exploitation de ces données ou le risque d'exploitation diminue l'intensité de l'usage de la consommation du service, et donc les revenus directs de l'entreprise sur ce marché. Au contraire, laisser la propriété des données aux consommateurs permet de garantir une utilisation raisonnable de celle-ci, à un coût d'opportunité limité puisque leur valeur marchande est faible. Si, au contraire, la valeur marchande potentielle des données est élevée, il est important pour les entreprises de les exploiter monétairement, et de les traiter pour en extraire leur valeur maximale. Dans ce cas, les entreprises sont motivées à proposer des offres très avantageuses aux consommateurs, si elles ont la possibilité de valoriser les données ainsi générées.

Au final, pour les consommateurs comme pour les firmes, il est optimal de laisser la propriété des données aux consommateurs lorsque ces données n'ont pas trop de valeur marchande et au contraire d'allouer les droits de propriété aux entreprises lorsque la valeur des données est importante. Autrement dit,



lorsque les données ont peu de valeur marchande, les inefficacités liées aux sous-investissements des entreprises dans le processus de valorisation sont peu importantes au regard du bénéfice que les consommateurs retirent d'un meilleur contrôle de l'usage de leurs données. En revanche, lorsque les données ont une valeur marchande importante, accorder aux entreprises le droit d'exploiter les données des consommateurs les conduit à créer une valeur suffisante pour faire plus que compenser les consommateurs pour les désagréments liés à l'usage de leurs données.

Au-delà des questionnements sur pertinence d'une monétisation des données personnelles, ce travail laisse en suspens des questions importantes. Tout d'abord, la détermination de la valeur des données personnelles est complexe. Les données sont utilisées en interne (pour améliorer le service proposé), mais également en externe (pour faire du ciblage ou bien être revendues sur un marché) et déterminer la valeur d'une donnée pour une firme est une entreprise très spéculative. De plus, la valeur des données personnelles d'un agent ne dépend pas uniquement de cet agent. En effet, il existe des externalités informationnelles entre consommateurs (voir Choi et coll., 2019) si bien que les informations relatives à une personne peuvent parfois extraites des données personnelles d'autres personnes (amis, consommateurs avec les mêmes profils...). Enfin, les transactions économiques générant les données d'une personne peuvent impliquer plusieurs firmes à la fois ou bien les mêmes données peuvent être générées de façon parallèle par des transactions économiques indépendantes. La question de la multipropriété, ou de l'exclusivité de la propriété se pose alors, rendant encore un peu plus complexe la constitution d'un marché de la donnée. ■

Publié initialement le 28 octobre 2020



## Références

Choi, J.P., Jeon, D.-S., & Kim, B.-C., (2019). Privacy and data collection with information externalities. *Journal of Public Economics*, 173, 113-124.

Coase, R. (1960). The problem of Social Cost. *Journal of Law and Economics*, 4, 1-44.

Dosis, A., & Sand-Zantman, W. (2019). The Ownership of Data. mimeo TSE and ESSEC.

Duch-Brown, N., Martens, B., & Mueller-Langer, F. (2017). The Economics of Ownership, Access and Trade in Digital Data. JRC Digital Economy, Working Paper 2017- 01.

Federal Trade Commission (2014). Data brokers: A call for transparency and accountability. Report. TC, Washington DC.

# À L'ÈRE DE L'IA, AVONS-NOUS ENCORE BESOIN DE CHERCHEURS ?



**Viviana Fang He** est Professeure associée au département de Management. Elle étudie la collaboration dans trois contextes liés à l'innovation : les équipes de nouvelles entreprises, les projets de recherche et développement et les communautés de logiciels libres. Ses recherches, qui combinent de manière inédite les théories organisationnelles et psychologiques, ont été publiées dans des revues internationales de premier plan et ont remporté des subventions nationales compétitives.

Écrit avec Julia Smith, rédactrice en chef, ESSEC Knowledge

## Avons-nous encore besoin de chercheurs étant donné l'essor et la portée de l'intelligence artificielle ?

Nous savons que le machine learning peut identifier des relations complexes dans des ensembles de données massifs qui ne sont pas nécessairement identifiables à l'œil "nu" (humain). Nous savons également que l'intelligence artificielle pourra un jour prendre en charge de nombreuses fonctions humaines. La recherche sera-t-elle l'une d'entre elles ? Viviana Fang He de l'ESSEC Business School et ses collègues Yash Raj Shreshta (ETH Zurich), Phanish Puranam (INSEAD), et Georg von Krogh (ETH Zurich) se penchent sur cette question dans leurs récentes recherches.

En un mot : non, mais l'histoire est un peu plus compliquée que cela.

Jusqu'à présent, les techniques du machine learning ont été largement utilisées pour coder des données et faire des prédictions, mais pas encore pour la tâche principale d'un chercheur : élaborer une théorie. Pourquoi en est-il ainsi ? Cela pourrait être dû à un

dégoût des chercheurs pour les soi-disant "prédictions sans explications". En fait, c'est exactement là que se situe l'opportunité, suggèrent le professeur He et ses collègues. Le machine learning pourrait en effet s'avérer meilleur que les chercheurs pour trouver des modèles solides et complexes dans les données.

Traditionnellement, les chercheurs en management proposent une théorie et/ou un modèle et le testent ensuite, généralement en utilisant un ensemble de données relativement petit. Avec des ensembles de données plus importants, il y a plus de chances que les résultats soient applicables à une population plus large plutôt qu'à celle utilisée dans l'étude et qu'ils soient donc reproductibles — ce qui est vrai dans d'autres situations que la présente. Les chercheurs peuvent également étudier davantage de variables lorsqu'ils travaillent avec des ensembles de données plus importants, ce qui est inestimable pour construire une image plus complète de la situation étudiée. Lorsqu'ils élaborent une théorie à partir de données en utilisant des outils statistiques traditionnels, les chercheurs courent le risque de surestimer les données, c'est-à-dire de trouver un modèle spécifique à l'échantillon actuel. Les algorithmes du machine learning



© Gettyimages - Igor Kuchev

ont des procédures qui aident à éviter le surajustement, ce qui signifie que les modèles qu'ils identifient ont plus de chances d'être reproduits dans d'autres échantillons. C'est une propriété très précieuse, car elle pourrait, par exemple, aider à résoudre la crise de répliation actuelle de la psychologie en facilitant l'élaboration de théories solides et donc de résultats reproductibles.

Un autre avantage de l'intégration des techniques du machine learning est qu'elle peut aider à gérer les préjugés des chercheurs en rendant les processus de recherche et les décisions transparents. Après tout, les chercheurs ne sont que des êtres humains. Il est donc possible qu'ils subissent un biais de confirmation et cherchent des résultats qui soutiennent leurs prévisions : en d'autres termes, qu'ils voient ce qu'ils veulent voir. En utilisant des algorithmes de machine learning, les chercheurs peuvent spécifier le niveau de complexité des modèles détectés et documenter ces décisions. L'ensemble de ces procédures permet une approche réfléchie pour équilibrer la précision des prédictions et la possibilité d'interprétation. Une précision prédictive plus élevée peut signifier que les modèles sont trop complexes pour être compris, et une interprétabilité plus élevée peut

signifier que le modèle est plus simple et ne tient peut-être pas compte de tous les facteurs d'impact. Il est essentiel de pouvoir contrôler ce compromis pour interpréter les modèles d'une manière qui ait un sens pour les personnes et pas seulement pour les machines. Cela signifie également que les chercheurs peuvent expliquer leur raisonnement de manière transparente.

Cependant, les machines ne peuvent pas agir seules : les algorithmes manquent d'intuition et du bon sens que possèdent les humains. S'ils peuvent assembler les pièces du puzzle, c'est à nous, les humains, d'expliquer pourquoi les pièces vont ensemble. De nombreuses parties essentielles du processus d'élaboration de la théorie resteront à la charge des chercheurs, comme la définition des facteurs d'intérêt, la sélection ou le développement de moyens de mesure de ces facteurs, et l'explication des relations qui sous-tendent les modèles observés. L'avenir de la théorisation nécessitera une synergie entre les algorithmes et les humains.

Le professeur He et ses collègues proposent une procédure en quatre étapes pour explorer cette possibilité. La première étape consiste à diviser

l'échantillon en deux : un échantillon à utiliser pour la détection de modèles assistée par le machine learning, et le second échantillon à utiliser pour tester les hypothèses. Lors de la deuxième étape, les chercheurs programment les algorithmes et ceux-ci font leur magie et identifient des modèles interprétables et fiables. À l'étape 3, les chercheurs se demandent si les modèles ont un sens et trouvent des explications à ces modèles. C'est à cette étape que l'expertise et le jugement humains sont essentiels, car les algorithmes du machine learning n'ont pas la capacité de le faire. À l'étape 4, les chercheurs testent les hypothèses — la théorie — dans le deuxième échantillon pour voir si le modèle se maintient.

Certains chercheurs ont déjà adopté cette approche, démontrant qu'elle peut être menée à bien. Elle a été utilisée pour étudier les conflits de gouvernance dans les communautés en ligne (He, Puranam, Shreshta, von Krogh, 2020), pour identifier le revenu optimal pour une large gamme de produits de l'App Store (Tidhar & Eisenhardt, 2020), et pour évaluer si une idée va ou non décoller (Dahlander, Fenger, Beretta, Kazami, & Frederiksen, 2020). Les approches similaires sont également expérimentées dans le domaine des sciences naturelles.



Par exemple, Udrescu et Tegmark (2020), deux physiciens du MIT, ont utilisé 100 équations pour générer des données puis les transmettre à un réseau de neurones. Leur algorithme a réussi à récupérer les 100 équations ! Cet ensemble diversifié d'études montre que l'approche peut être appliquée à une grande variété de sujets, ce qui la rend utile aux chercheurs de toutes les disciplines.

Bien que cette approche ait des implications importantes pour l'élaboration de théories, les auteurs notent qu'il faut tenir compte de certaines réserves avant de l'utiliser.

Le machine learning suppose que l'avenir peut être prédit à partir du passé, il est donc préférable d'utiliser des algorithmes lors de l'évaluation de phénomènes relativement stables. Deuxièmement, il ne peut pas remplacer la randomisation. Les techniques de machine learning sont plus adaptées pour faire des prédictions que pour tester une théorie sur les relations entre les variables. Il existe également le risque que les techniques du machine learning amplifient les biais présents dans les données, ce qui conduirait à des conclusions biaisées, car les biais pourraient être difficiles à détecter, mais avoir des conséquences éthiques importantes. C'est pourquoi les chercheurs doivent avoir une solide compréhension conceptuelle des techniques qu'ils utilisent, ce qui n'est pas chose facile dans un domaine qui progresse si rapidement.

En bref, si le machine learning ne peut pas remplacer les chercheurs, il peut prendre en charge certaines fonctions que les humains effectuent actuellement, comme la reconnaissance des formes, la mémorisation par cœur et l'arithmétique. Cependant, les êtres humains sont nécessaires pour les tâches qui requièrent plus d'intuition et de créativité, comme l'explication de modèles, la rédaction de livres et l'art.

Alors, avons-nous encore besoin de chercheurs ? Oui — et le machine learning peut être un outil puissant pour produire des recherches plus solides. ■

Pour plus d'informations sur l'utilisation des algorithmes du machine learning dans la construction de théories, consultez leur article ici. <https://pubsonline.informs.org/doi/full/10.1287/orsc.2020.1382>

Publié initialement le 4 janvier 2021

## Références

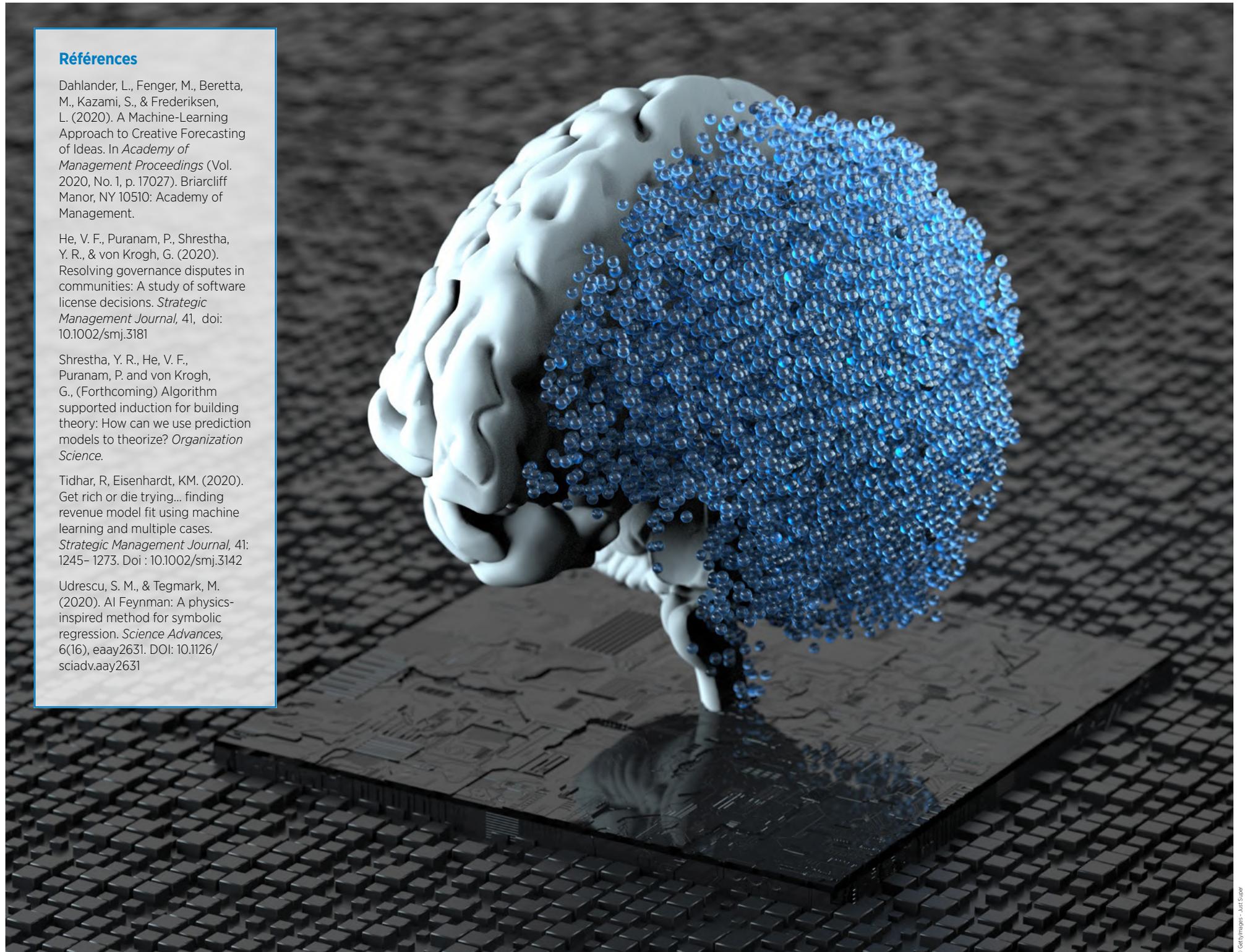
Dahlander, L., Fenger, M., Beretta, M., Kazami, S., & Frederiksen, L. (2020). A Machine-Learning Approach to Creative Forecasting of Ideas. In *Academy of Management Proceedings* (Vol. 2020, No. 1, p. 17027). Briarcliff Manor, NY 10510: Academy of Management.

He, V. F., Puranam, P., Shrestha, Y. R., & von Krogh, G. (2020). Resolving governance disputes in communities: A study of software license decisions. *Strategic Management Journal*, 41, doi: 10.1002/smj.3181

Shrestha, Y. R., He, V. F., Puranam, P. and von Krogh, G., (Forthcoming) Algorithm supported induction for building theory: How can we use prediction models to theorize? *Organization Science*.

Tidhar, R., Eisenhardt, KM. (2020). Get rich or die trying... finding revenue model fit using machine learning and multiple cases. *Strategic Management Journal*, 41: 1245– 1273. Doi : 10.1002/smj.3142

Udrescu, S. M., & Tegmark, M. (2020). AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16), eaay2631. DOI: 10.1126/sciadv.aay2631



# IA ET GOUVERNANCE ? DE L'INFORMATION UNE RENCONTRE PROMETTEUSE

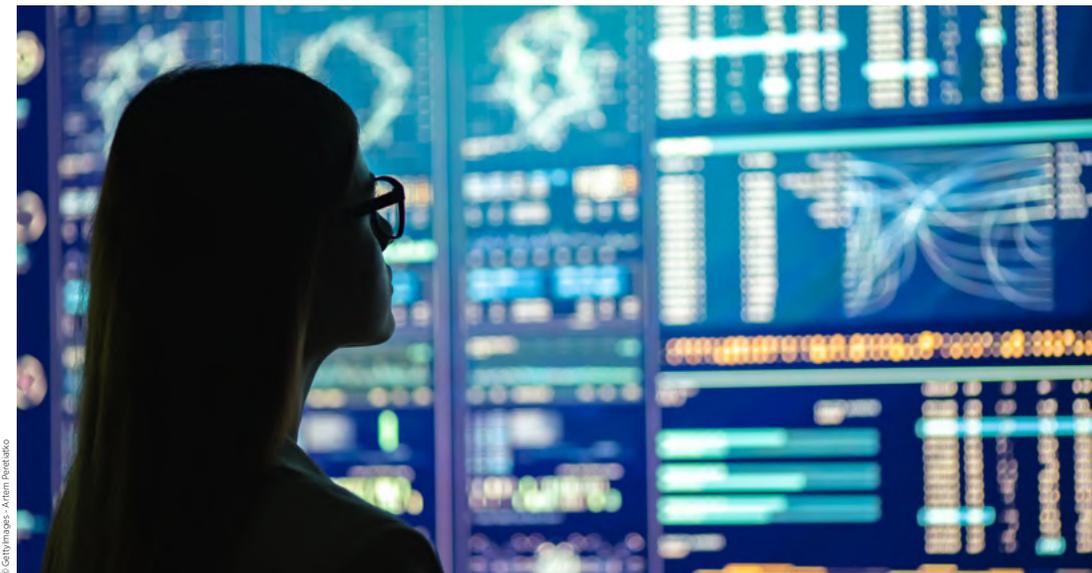


**Isabelle Comyn-Wattiau** est Professeure à l'ESSEC Business School où elle enseigne les systèmes d'information. Elle a fondé la chaire *Information Strategy and Governance*. Ses projets de recherche actuels portent sur la stratégie et la gouvernance de l'information, les sciences humaines numériques, la modélisation et la conception des systèmes d'information. Ses recherches ont été publiées dans de nombreuses revues universitaires de premier plan, notamment le *Journal of Management Information Systems*, *Decision Support Systems*, *Expert Systems with Applications*, *Data and Knowledge Engineering*, et quelques autres. Elle préside le comité directeur de la conférence internationale sur la modélisation conceptuelle. Elle fait partie du comité de rédaction de plusieurs revues universitaires, notamment *Data and Knowledge Engineering* et *Systèmes d'information et gestion*. Elle a été rédactrice en chef de la revue *Ingénierie des Systèmes d'Information (Lavoisier)* et a récemment créé la revue ouverte *D'Ingénierie des Systèmes d'Information*.

L'intelligence artificielle a envahi le monde des affaires, annonçant des bouleversements dans l'aide à la décision. C'est une nouvelle brique dans l'ensemble des composants informationnels que toute organisation doit intégrer ou, à tout le moins, prendre en compte. La mise en place du règlement général sur la protection des données (RGPD) en Europe en 2018 et ses équivalents dans les autres pays ou continents ont mis en lumière la difficulté et l'enjeu d'une bonne gouvernance de l'information au sein de toutes les organisations. Dans cet article, nous confrontons ces deux éléments de la nouvelle donne : l'intelligence artificielle et la gouvernance de l'information. Après avoir rappelé ce qu'est cette gouvernance, nous aborderons les deux facettes de la confrontation : comment la gouvernance de l'information doit prendre en compte les outils d'intelligence artificielle, consommateurs et producteurs d'une information à gouverner ? Comment ces mêmes outils peuvent-ils aider l'organisation dans sa gouvernance des données et de l'information ?

## 1. Qu'est-ce que la gouvernance de l'information ?

La gouvernance de l'information désigne la mise en place, au sein d'une entreprise ou d'une organisation, d'une politique de management de l'information, avec des objectifs définis et une mise en œuvre planifiée, incluant des moyens humains, organisationnels et technologiques. Les termes de gouvernance des données et gouvernance de l'information sont utilisés, souvent comme synonymes ou variantes, pour désigner cette stratégie de pilotage des actifs informationnels. Préférer l'expression gouvernance des données permet d'attirer l'attention sur l'aspect digital du sujet, mais présente le risque de limiter la réflexion aux données purement structurées. La gouvernance de l'information ou gouvernance des informations englobe elle aussi les données, quelle que soit leur forme, données structurées dans des tables, données semi-structurées dans des documents ou données moins structurées dans les logiciels de messagerie ou sous forme d'image. Finalement, la gouvernance de l'information peut être définie par son objectif : maximiser la valeur de l'information en minimisant les coûts et les risques associés.



© Gettyimages - Artem Pevetako

La gouvernance de l'information a en charge un ensemble de processus-clés, parmi lesquels le management de la qualité des données ou encore leur sécurité (disponibilité, intégrité, confidentialité, mais aussi traçabilité). Même s'ils s'appuient sur des solutions informatiques, ces processus sont souvent, à juste titre, sous la responsabilité des métiers.

La gouvernance de l'information est pertinente dans tous les domaines et secteurs d'activité. Elle prend plus d'acuité là où les trois dimensions complémentaires de valeur, coût et risque sont prégnantes. Prenons par exemple le secteur de la banque où le coût de gestion de l'information peut représenter plus de 10 % du chiffre d'affaires et le risque associé est important du fait par exemple de la manipulation d'une information très particulière qu'est l'argent sous forme électronique ! Dans le secteur de la santé, c'est la dimension risque — de divulgation ou d'erreur — qui prime, sans pour autant que l'on puisse occulter les autres dimensions.

La gouvernance de l'information s'organise de mieux en mieux dans les entreprises publiques et privées, et plus

largement dans toutes les organisations. Ces dernières montent en maturité sur le sujet, comme l'indiquent certaines études. Pour autant, elle doit faire face constamment à de nouveaux défis, l'intelligence artificielle en est un.

## 2. L'intelligence artificielle, une information bien particulière à gouverner

Pour comprendre, décider, apprendre, les outils d'intelligence artificielle emmagasinent des données ou des informations. Ces dernières sont de plus en plus volumineuses et variées. Elles doivent, comme toutes les informations de l'entreprise, être sous contrôle, et c'est un des aspects importants de la gouvernance de l'information. Qu'il s'agisse d'entrepôts de données, de lacs de données, ou tout autre réservoir d'information, rien ne doit échapper à la cartographie de l'information et s'inscrire dans les trois dimensions de valorisation, de gestion des risques et de contrôle des coûts.

La qualité des données prend tout son sens : elle doit faciliter l'utilisation de l'intelligence artificielle, en mettant à

disposition des processus de collecte, de repérage et de nettoyage des données. Au-delà de ces aspects classiques de la gouvernance des données, l'intelligence artificielle soulève de nouvelles questions comme celle des biais humains présents tant dans les algorithmes que dans les données et qui peuvent conduire à de mauvaises décisions. Outre le repérage des données personnelles et, plus généralement, des données sensibles, le référencement des informations issues de l'intelligence artificielle doit aussi intégrer une description du fonctionnement de ces algorithmes.

C'est tout l'enjeu de l'IA explicable (en anglais, explainable AI) qui désigne les méthodes permettant d'équiper les boîtes noires de l'intelligence artificielle de modules expliquant, au moins partiellement, les résultats fournis. C'est un enjeu réglementaire dans de nombreux cas, c'est aussi un enjeu de confiance afin de permettre aux utilisateurs des informations ou conclusions de l'intelligence artificielle de s'appuyer sans réticence sur cette aide précieuse. Enfin, c'est un enjeu d'éthique pour les organisations qui sont ainsi en mesure d'assurer une transparence de leurs processus, même quand ceux-ci ont recours à des outils d'intelligence artificielle.





## L'intelligence artificielle au service de la gouvernance de l'information

À l'inverse, les outils d'intelligence artificielle peuvent aussi se révéler une aide précieuse pour la gouvernance de l'information.

Ainsi, la cartographie de l'information doit, entre autres, repérer partout les données personnelles pour assurer

la conformité au RGPD. L'intelligence artificielle peut aider à relever le défi de la maintenance de cette cartographie, tâche fastidieuse, mais non triviale, qui requiert des capacités d'analyse et de prise de décision. Le volume et l'omniprésence des données rendent cette tâche impossible à assurer manuellement.

L'homme va rester maître de cet écosystème qu'est la gouvernance de l'information, mais l'intelligence artificielle et d'autres technologies

émergentes, comme la blockchain, vont jouer un rôle très important d'assistance pour analyser les informations et assurer la conformité. L'automatisation peut concerner aussi l'archivage et la destruction de l'information, phase ultime du cycle de vie, là encore pour répondre tant aux obligations réglementaires qu'aux besoins d'efficacité.

Les éditeurs de logiciels spécialisés dans la gouvernance des données disent avoir intégré l'intelligence artificielle

dans leur palette de composants même s'il reste difficile d'en évaluer la portée. Certains utilisent des techniques intelligentes pour repérer les nouvelles données, identifier les liens entre elles, classer les documents. Les techniques d'apprentissage permettent de décrire des contenus types que le logiciel devient capable de reconnaître après un entraînement suffisant. Parmi les techniques relevant de l'intelligence artificielle, le traitement du langage naturel, la reconnaissance automatique de caractères continuent à s'améliorer

de plus en plus. Les robots logiciels exécutent des tâches répétitives, repèrent par exemple des factures dans un ensemble de documents administratifs, fournissant ensuite à l'expert humain une analyse qu'il n'a plus qu'à confirmer, à l'image du chèque traité automatiquement dans tous les guichets automatiques des banques.

Là aussi, une intelligence artificielle explicable sera de nature à aider l'organisation à faire la preuve de sa bonne gestion de l'information pour

faire face à la réglementation ou pour instruire une demande de vérification. Finalement, face à la réglementation qui évolue elle aussi et qui varie d'un pays à un autre, en présence d'informations de plus en plus riches et variées, la gouvernance de l'information a aussi beaucoup à gagner en recourant aux outils de l'intelligence artificielle. ■

*Publié initialement le 16 décembre 2020*



# IA AU SERVICE DE L'OCTROI DES CRÉDITS AUX PARTICULIERS: PROMESSES ET DÉFIS



**Andras Fulop** est Professeur au département de Finance. Il enseigne la fixation empirique des prix des actifs, les salaires fixes et la gestion des risques financiers dans les programmes PhD, master en finance et master en gestion. Ses recherches portent globalement sur le prix des actifs, notamment pour les risques de crédit, l'économétrie financière et les produits dérivés. Andras a publié dans des revues académiques telles que la *Review of Financial Studies*, le *Journal of Econometrics* et le *Journal of Business and Economic Statistics*. Il est docteur en finance de la Rotman School of Management de l'Université de Toronto.



**Laurence Daures-Lescourret** est Professeure associée au département de Finance. Ses recherches portent sur la microstructure des marchés financiers. Laurence a reçu le prix de la thèse de doctorat en 2004 de l'Association française de finance et de la Fondation nationale de l'enseignement supérieur en gestion d'entreprise (FNEGE), le "Prix Joseph de la Vega" en 2013 et le prix IFSID pour le meilleur article sur les produits dérivés en 2015. Elle a reçu plusieurs bourses de recherche d'Euronext Paris (2007), du FEI (2008, 2010, 2020), de l'Agence nationale française de la recherche (JCJC, 2011) et de l'INEX (Initiative d'Excellence, 2018). Elle siège au conseil d'administration de Dassault Systèmes et de LCL (Le Crédit Lyonnais). Elle siège au comité d'audit de Dassault Systèmes (DS) et préside le comité de rémunération et de nomination de DS. Elle siège au comité des risques de LCL et préside le comité d'audit de LCL. Laurence est titulaire d'un doctorat en finance de HEC Paris.

Les technologies de l'information impactent le risque de crédit depuis plusieurs dizaines d'années. La Fair Isaac Corporation (FICO) a introduit les « scores » FICO en 1989, initiant une transition entre une analyse de solvabilité des crédits réalisée par des opérateurs humains et celle automatisée par des algorithmes. Le « score » FICO est une note de crédit établissant le risque de solvabilité de l'emprunteur, qui s'appuie sur cinq éléments: l'historique des paiements, le niveau d'endettement courant, les types de crédit détenus, l'ancienneté de l'historique de crédit et les nouveaux crédits. Aujourd'hui, ce score est utilisé dans plus de 90 % des décisions de crédit prises aux États-Unis.

Avec la digitalisation croissante de la société au cours de la dernière décennie, il y a eu une explosion à la fois dans la collecte de données personnelles, dans la sophistication des algorithmes et dans la capacité de calcul de traitement des informations. Cette accélération digitale a le potentiel d'automatiser totalement le processus d'évaluation de solvabilité des individus. Dans cet article, nous résumons les principales leçons extraites de la littérature académique récente concernant l'application des techniques d'intelligence artificielle (IA) à l'octroi de crédit.



## Comment l'IA peut-elle aider la décision d'octroi de crédit ?

Les nouveaux outils d'IA (machine learning, deep learning) sont très prometteurs. Ils améliorent significativement les techniques existantes de détermination du risque de crédit des emprunteurs. Pour comprendre les voies d'amélioration, rappelons que les modèles traditionnels de notation du crédit, comme, par exemple, celui du score FICO, utilisent des sources d'informations financières issues des comptes bancaires des particuliers.

Ensuite, un nombre relativement limité de variables prédictives est généralement combiné au sein de modèles linéaires de la probabilité de défaut de l'emprunteur (tels que les régressions logistiques) afin d'en déduire la note de crédit. Les nouvelles approches d'IA vont au-delà de ces outils sur au moins deux aspects.

Premièrement, les méthodes de « machine learning » (ou apprentissage automatique), telles celles à arbre de décision ou de réseaux neuronaux, peuvent faire de la prédiction à partir

de relations non linéaires entre les variables prédictives et le risque de crédit individuel, tout en permettant d'éviter les écueils de l'« overfitting » (lorsque le modèle, en essayant de trop coller aux données, n'est plus généralisable). Généralement, les techniques de machine learning surpassent les modèles de prédiction linéaires traditionnels, en particulier, pour les groupes d'emprunteurs à risque élevé. Walther et al (2020), en utilisant des données relatives à des millions de prêts immobiliers américains, ont montré que les méthodes d'arbres de décision surpassent considérablement les techniques de régression logistique. Albanesi et Vamossy (2019) ont comparé plusieurs algorithmes de machine learning sur les données de crédit de l'entreprise Experian spécialisée en risque de crédit et ont constaté qu'un ensemble de techniques combinant réseaux de neurones et arbres améliore les modèles traditionnels et que cette amélioration est particulièrement prononcée pour les emprunteurs avec des notations de crédit dégradées.

Deuxièmement, la digitalisation de notre mode de vie et l'intelligence artificielle permettent d'utiliser de nouveaux types de données. En

utilisant les données de connexion relatives à environ 270 000 achats en ligne, Berg et al. (2020) ont analysé le pouvoir prédictif des « empreintes digitales » (les informations laissées par les individus lors de la visite d'un site Web, comme le type de système d'exploitation utilisé, l'adresse e-mail, l'heure d'achat, etc.). Ils ont constaté que la précision d'un modèle de risque de crédit utilisant des variables d'empreinte digitale est comparable et complémentaire à celle des scores de risque de crédit traditionnels. Par conséquent les empreintes digitales peuvent être des outils d'aide à la décision intéressants, en particulier pour établir le profil d'emprunteurs potentiels qui n'auraient pas assez d'historique de crédit pour obtenir une notation fiable donnant accès au crédit. Les données non structurées sous forme de texte ou d'images (par exemple, des informations provenant de réseaux sociaux tels que LinkedIn, Twitter ou Facebook) constituent une autre source d'informations utilisée par les algorithmes d'IA. En utilisant les données de Prosper, une plateforme de financement participatif, Netzer et al (2019) constatent que le fait de compléter les informations financières et démographiques par les informations





**Pour aller plus loin**

Albanesi, S., & Vamossy, D. F. (2019). Predicting consumer default: A deep learning approach (No. w26165). *National Bureau of Economic Research*.

Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2019). Consumer-lending discrimination in the FinTech era (No. w25943). *National Bureau of Economic Research*.

Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the rise of fintechs: Credit scoring using digital footprints. *The Review of Financial Studies*, 33 (7), 2845-2897.

Iyer, R., Khwaja, A. I., Luttmer, E. F., & Shue, K. (2016). Screening peers softly: Inferring the quality of small borrowers. *Management Science*, 62 (6), 1554-1577.

Netzer, O., Lemaire, A., & Herzenstein, M. (2019). When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research*, 56 (6), 960-980.

Walther, A., Ramadorai, T., Goldsmith-Pinkham, P., et al., (2020). Predictably unequal? The effect of machine learning on credit markets, *The Journal of Finance*, forthcoming.

© Gettyimages - Wladimir

textuelles soumises par les emprunteurs potentiels améliore considérablement la prédiction des défauts.

### Quelles sont les implications du recours à l'IA pour les consommateurs ?

Sur une note positive, les améliorations de la prédiction du risque de solvabilité sont particulièrement prononcées parmi les groupes plus risqués et les personnes ayant des antécédents de crédit limités (Berg et al, 2020). L'IA permet donc de réduire les problèmes d'information asymétrique entre emprunteurs et

créanciers et donc de développer l'accès au crédit. Cela peut être particulièrement utile pour les pays émergents dans lesquels le secteur bancaire est souvent embryonnaire et donc dans lesquels la plupart des consommateurs sont dépourvus d'informations financières traditionnelles utilisées dans les modèles de crédit standard freinant ainsi l'accès et le développement du crédit.

Un autre avantage potentiel du transfert des décisions de crédit aux algorithmes est que les biais humains - tels que le racisme ou les préjugés contre les minorités - peuvent être court-circuités (au moins théoriquement), conduisant à moins de discrimination.

Pour examiner cette question de manière empirique, il faut d'abord une définition précise de la discrimination. Bartlett et al (2019) suggèrent d'utiliser l'interprétation des tribunaux américains, selon laquelle tout impact différentiel de traitement des groupes minoritaires non lié à la « nécessité commerciale légitime » est jugé discriminatoire. Bartlett et al (2019) constatent que, concernant la tarification des prêts immobiliers, les algorithmes FinTech discriminent également, mais 40 % de moins que les créanciers humains. En revanche, les FinTechs ne font pas de discrimination à la décision d'octroi de ces prêts, contrairement aux créanciers humains.

### Les défis des algorithmes d'IA

Si promesses il y a, l'IA au service de l'octroi du crédit fait également face à de sérieux défis. Premièrement, une précision plus forte du risque de solvabilité par les algorithmes amène à une plus grande disparité, voire inégalité, du coût du crédit entre consommateurs. Cette augmentation de la dispersion est particulièrement prononcée pour les minorités et les emprunteurs à risque plus élevé (Fuster et al, 2020). La prise en compte par les régulateurs de cette limite est nécessaire pour que les conditions de crédit des ménages défavorisés soient améliorées et moins inégalitaires.

Deuxièmement, l'utilisation des nouvelles données (par exemple, l'université ou grande école dont est diplômé le potentiel emprunteur) par les algorithmes d'IA peut potentiellement créer des discriminations par inadvertance, comme le soulignent Bartlett et al (2019). Prenons l'exemple d'une entreprise qui met en place un algorithme de filtrage des emprunteurs pour maximiser ses profits, mais que ce filtrage introduise ex post un impact différentiel sur les groupes minoritaires protégés. Par exemple l'algorithme prend en compte la donnée relative à l'éducation de l'emprunteur, et donc l'université ou grande école de ce dernier, qui est certes un proxy de la

richesse de l'emprunteur, mais qui est probablement corrélée avec sa race, ou son ethnie. L'entreprise risque de tomber sous le coup d'une législation anti-discriminatoire, même si l'algorithme ne contient pas de biais contre les minorités. La nature de la boîte noire de la plupart des algorithmes d'IA augmente ce risque de discrimination par inadvertance.

Au final, l'avenir des algorithmes d'IA dans les décisions d'octroi de crédit est réellement prometteur, mais ses opérateurs humains doivent veiller à réduire le risque de décisions inéquitables pour fournir des évaluations justes et précises pour tous. ■  
Publié initialement le 18 décembre 2020



# COMMENT LES DONNÉES ET LES TALENTS DU DATA SONT ESSENTIELS POUR LE SUCCÈS DU DÉVELOPPEMENT DES ÉNERGIES RENOUVELABLES



**Jeroen Rombouts** est Professeur au département de Statistiques et d'économétrie à l'ESSEC Business School depuis 2013. Il est le titulaire de la chaire Accenture Strategic Business Analytics. Il a obtenu son doctorat en économétrie en 2004. Il donne des cours d'analyse de données & de science des données aux programmes MBA et masters à Paris & à Singapour. Il étudie l'analyse prédictive, en particulier l'analyse des séries chronologiques et les prévisions. Il a publié de nombreux articles scientifiques, organise des ateliers et est rédacteur en chef adjoint de plusieurs revues scientifiques quantitatives. Ses recherches sont financées par la Commission européenne et l'Agence nationale de la recherche. Il intervient fréquemment en tant que consultant expert sur des sujets d'analyse stratégique d'entreprise et de science des données. Avant de rejoindre l'ESSEC, il était professeur associé à HEC Montréal.

**Co-auteur**  
**Gérard Guinand** est le Group Chief Data Officer à Engie.

**A**u cours des dernières années, les GAFAM ont montré que l'utilisation de données pour connaître ses clients est essentielle pour le développement de nouveaux produits et services et pour battre la concurrence. Ces entreprises mettent le digital au cœur de leur fonctionnement, permettant une prise de décision basée sur les données et enrichie par celles-ci, et sont très appréciées des investisseurs étant donné leur capitalisation boursière massive. Même pendant la pandémie COVID-19, le secteur de la technologie a continué à croître de manière spectaculaire grâce à l'accélération de la numérisation de notre façon de travailler et d'interagir avec les autres.

Toute grande entreprise mature s'inspire de la façon dont les entreprises de technologie fonctionnent et dominant. Par exemple, en France, L'Oréal veut devenir la meilleure entreprise de « Beauty Tech » en utilisant l'intelligence artificielle et la réalité augmentée. Depuis 2018, Carrefour a utilisé le Carrefour-Google Lab pour accélérer sa transformation numérique. En 2020, Danone et Microsoft ont lancé « The AI Factory

for Agrifood » (l'usine d'intelligence artificielle pour l'agroalimentaire). Les entreprises de l'énergie comme Engie et EDF sont poussées par l'opinion publique sur le changement climatique à devenir opérationnellement excellentes et plus vertes. De jeunes talents des données sont engagés pour aider à transformer les entreprises et à introduire la nouvelle culture des données.

## Mentalité prédictive

En théorie, l'utilisation intelligente des données et la création d'une valeur commerciale ont beaucoup de sens, bien qu'en pratique, les entreprises traditionnelles aient du mal à s'orienter davantage vers les données. Ces dernières années, les entreprises ont largement investi dans l'infrastructure de données, nommé des responsables des données et lancé des programmes de formation pour convaincre chaque salarié de l'importance des données et de l'analyse. Par conséquent, des quantités massives de données sont stockées dans le nuage et la question qui se pose souvent est « que pouvons-nous faire avec cela »



ou encore « quel est le retour sur tous ces investissements en matière de données ». Pour répondre à ces questions, il est essentiel de passer à l'étape suivante du processus de maturité des données de l'entreprise. C'est l'étape qui permet de devenir plus informé, plus axé et meilleur sur le plan opérationnel, et cela nécessite d'utiliser les données pour regarder vers l'avenir, plutôt que vers le passé, et donc faire des prévisions. En d'autres termes, après avoir stocké et classé les données, il est temps de les utiliser pour la prise de décision à tous les niveaux de l'organisation, plutôt que dans des secteurs spécifiques. En fait, les décisions de l'entreprise comportent toujours implicitement des prédictions, et il est temps de rendre ce processus plus formel et automatique grâce à l'utilisation des données.

Le paradigme prédictif ne concerne pas seulement les algorithmes de recommandation et autres: il permet également d'utiliser les données au niveau exécutif pour s'assurer que la stratégie est mise en œuvre. Plus précisément, la gestion d'indicateurs de performance clés (KPI) prospectifs permet de mesurer et de suivre le

succès de l'entreprise et de mettre en place des objectifs clairs. Cela génère des données précieuses qui peuvent être corrélées avec de nouvelles initiatives pour prédire leur succès et mieux comprendre leur lien avec les opérations existantes. Pour résumer, les cadres supérieurs doivent commencer à mettre en œuvre une stratégie avec des données plutôt qu'une stratégie pour des données, afin que le modèle d'exploitation de l'entreprise puisse devenir centré sur les données, comme le sont les célèbres entreprises de technologie telles qu'Amazon et Ali Baba.

Quelqu'un a dit un jour: « Faire des prédictions est difficile, surtout en ce qui concerne l'avenir ». Les prédictions sont par nature incertaines et il faut en tenir compte lors de la prise de décisions commerciales, à l'instar des investisseurs financiers qui utilisent plus d'informations que le rendement moyen d'un actif pour décider de l'acheter ou de le vendre. Des prévisions précises sont obtenues en combinant différentes sortes de données, y compris des sources externes comme les données météorologiques dans les applications énergétiques. La

production de prévisions à partir de données n'est pas une tâche triviale. Elle exige des Data scientists talentueux et des algorithmes avancés, ainsi qu'un contrôle continu des performances. Il n'est pas surprenant que l'International Data Corporation s'attende à ce que les dépenses mondiales en matière d'intelligence artificielle passent de 43 milliards d'euros à 94 milliards d'ici 2024.

## L'énergie renouvelable

L'Agence internationale de l'énergie projette que les énergies renouvelables fourniront 80 % de la croissance de la demande mondiale d'électricité d'ici 2030. En fait, les projets d'énergie solaire et éolienne sont devenus moins coûteux et les taux d'intérêt sont aujourd'hui historiquement bas. En outre, les gouvernements sont très favorables, comme en témoigne le « Green Deal » européen, dont la principale ambition est de rendre l'UE neutre sur le plan climatique d'ici 2050. Les énergies renouvelables deviennent donc un objectif stratégique clé pour les entreprises du secteur de l'énergie. Techniquement parlant,



les énergies renouvelables (en particulier l'éolien et le solaire) ont un niveau d'intermittence élevé (nuit, absence de vent) mais nous ne pouvons pas augmenter le nombre de centrales pour compenser ce manque de production pour des raisons économiques et environnementales. Cette situation implique deux actions principales pour les entreprises du secteur de l'énergie telles qu'ENGIE. Premièrement, identifier les meilleurs sites pour une nouvelle implantation. Deuxièmement, obtenir les meilleures performances des centrales tout en tenant compte des contraintes d'exploitation (le bruit par exemple) pour un volume plus important d'électricité produite.

Les données jouent un rôle essentiel pour le succès du développement des énergies renouvelables, car elles permettent la meilleure sélection des sites optimaux en fonction de la topographie et des données de prévisions météorologiques, les meilleurs résultats en fonction de la disponibilité technique, des données météorologiques en temps réel et des données de mesure liées aux contraintes environnementales (bruit), et l'optimisation des ventes d'électricité en croisant les données de production avec les données sur la demande, les prix du marché et les capacités de stockage.

En termes de retour sur investissement des données, les énergies renouvelables constituent un cas d'utilisation parfait où la valeur des données peut être rendue explicite. En effet, ces sources d'énergie sont hautement équipées de capteurs permettant une maintenance prédictive, mais aussi une monétisation des données. Du côté de la production, il n'y a pas de problème de GDPR. L'excellence opérationnelle dans le domaine des énergies renouvelables sera la seule façon de survivre pour les opérateurs historiques. Les compagnies

pétrolières traditionnelles telles que B.P. et Total se transforment rapidement et vont se trouver en concurrence féroce avec les acteurs actuels du marché des énergies renouvelables.

Engie et l'ESSEC Business School travaillent depuis trois ans sur différentes études de cas dans le cadre de la chaire de Strategic Business Analytics en partenariat avec Accenture. L'objectif principal de la Chaire est de former la prochaine génération de dirigeants à développer de nouvelles stratégies commerciales, en tirant parti des nombreuses applications de l'analyse avancée. Grâce à une méthode d'apprentissage hybride basée sur l'innovation, la collaboration et l'esprit d'entreprise, la chaire agit comme le cœur d'un écosystème combinant données et création de valeur — de l'élaboration d'objectifs et de stratégies à la transformation, en passant par la résolution de problèmes, la science des données et l'intelligence artificielle, le changement de culture et le développement des compétences.

Engie est une partie importante de l'écosystème de la chaire Strategic Business Analytics. En 2021, les étudiants de la chaire travailleront sur deux cas stratégiques relatifs aux énergies renouvelables. Leurs idées et perspectives axées sur l'avenir sur ces sujets génèrent des idées innovantes et des solutions de valeur. Les étudiants de l'ESSEC sont particulièrement intéressés à travailler avec des entreprises comme Engie, étant donné ses valeurs stratégiques fortement orientées vers l'écologie. En effet, préoccupés par le changement climatique, ils préfèrent travailler sur des études de cas qui génèrent en fin de compte une valeur sociétale plutôt que sur des cas purement commerciaux pour les plateformes d'e-commerce.

## L'avenir

À l'avenir, les énergies renouvelables créeront de nombreux emplois nécessitant des compétences techniques, informatiques et analytiques. Selon l'UE, l'industrie solaire photovoltaïque représente à elle seule 81 000 emplois, et devrait passer à 175 000 et 200 000-300 000 emplois en 2021 et 2030 respectivement. La numérisation et les énergies renouvelables vont de pair et constitueront un important moteur de croissance économique. Le partenariat de l'Engie et l'ESSEC garantira que les jeunes talents sont formés et acquièrent les compétences nécessaires pour assurer la transition vers une société verte en satisfaisant aux accords sur le climat.

Plus généralement, ce sont les entreprises qui emploient les personnes ayant les compétences, l'état d'esprit et la vision appropriés qui feront la différence. Les données sont désormais disponibles, la plupart des outils d'analyse qui créent de la valeur sont standards et les ressources informatiques sont peu contraignantes. C'est la culture de l'entreprise qui nécessite un changement fondamental. Ceux qui seront capables d'attirer de jeunes diplômés en management « prêts pour les données » seront à l'avant-garde de la concurrence.

Dans sa Vision technologique 2020, Accenture mentionne le choc technologique comme étant une nouvelle situation dans laquelle les individus sont d'une part enthousiastes à l'égard de la technologie, des données et de l'intelligence artificielle, tout en ayant besoin que les algorithmes soient compréhensibles et équitables, et qu'ils sachent où leurs données personnelles seront utilisées. Cet équilibre sera extrêmement important dans l'ère d'après-crise, où tout ce qui compte sera l'expérience humaine. ■

Publié initialement le 30 novembre 2020



# L'INTELLIGENCE ARTIFICIELLE ET ÉTHIQUE



**Laurent Bibard.** Docteur en philosophie et en économie, habilité à diriger des recherches en gestion et en philosophie, Laurent Bibard est Professeur au département de Management de l'ESSEC, où il enseigne la philosophie politique, la sociologie, et l'économie. Laurent a été à la tête de la Chaire Edgar Morin sur la complexité de 2014 à 2019. Il s'intéresse particulièrement aux dynamiques de vigilance en situation de crise et aux relations entre les sexes.

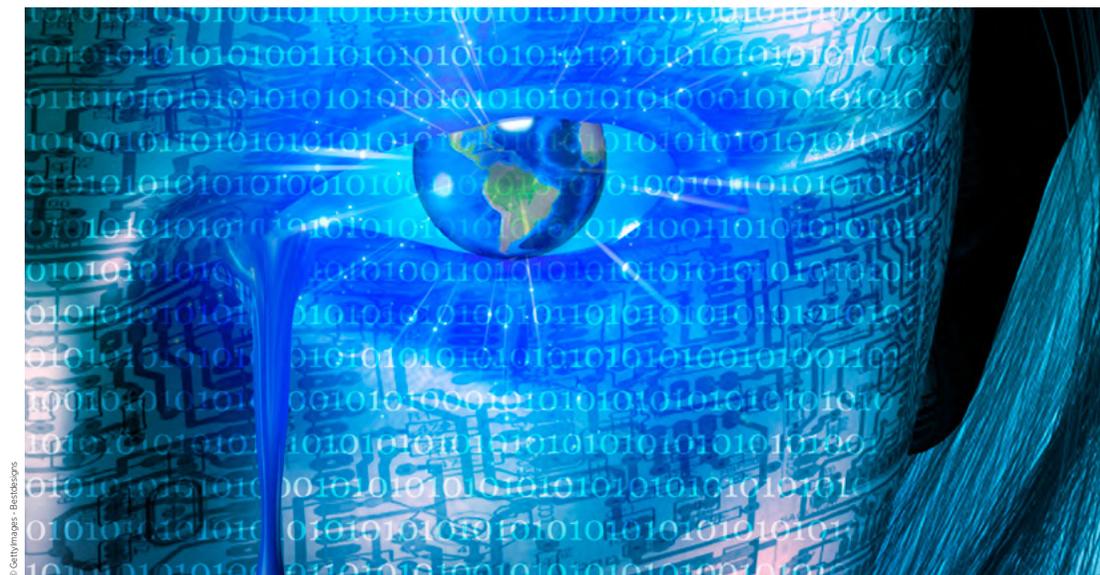
**A**u sujet des questions éthiques posées par l'intelligence artificielle, deux points majeurs doivent être soulignés :

- 1) Il n'y a pas de problème éthique au développement des nouvelles technologies et de l'intelligence artificielle en tant que telles : ce ne sont pas les technologies par elles-mêmes qui posent problème, mais l'usage que nous en faisons et que nous rêvons d'en faire. Vieille comme le monde, cette observation au sujet des « techniques » est capitale pour poser correctement et contre nos naïvetés, le problème de l'« éthique » concernant l'intelligence artificielle.
- 2) Les manières dont se pose de nos jours le problème sont en quantité et en variété infinies. Elles impliquent en particulier la fascination que provoquent les « nouvelles technologies » et les progrès de l'intelligence artificielle en termes de performance. Le problème le plus grave est actuellement qu'il est en général présumé et tacitement ou explicitement admis que l'intelligence artificielle non seulement va dominer l'intelligence humaine dans l'avenir mais qu'elle le

fait déjà. Cette présupposition prend la forme, chez les concepteurs de systèmes, de la présupposition que les humains ne sont ultimement que sources de défaillances. Une des conséquences majeures de cette présupposition est que les ingénieurs qui conçoivent les systèmes censés remplacer les humains sont de facto et sans contestation possible, bien supérieurs en tout – dont en éthique... – aux humains censés piloter les dits systèmes. Ceci se révèle particulièrement lorsque l'on est censé anticiper les problèmes « éthiques » des décisions en situation d'urgence comme pour les futurs véhicules dits « autonomes », ou encore dans l'aviation.

Je m'arrête au cas de l'aviation, particulièrement révélateur d'une posture profondément délétère non seulement en termes d'éthique, mais en termes de management, et de rapport à autrui de façon générale.

Outre les problèmes éthiques gravissimes liés au court-termisme et à la recherche d'économies dans lesquels Boeing a par exemple été pris lors de la conception précipitée du Boeing 737 MAX, se pose un problème



éthique spécifiquement lié à la conception, la mise en place et les (dys-) fonctionnements du système MCAS censé assurer la sécurité des avions en cas de décrochage pour perte de vitesse.

Le système MCAS est censé provoquer, lorsqu'un avion est en perte de vitesse et risque de décrocher, la plongée de l'appareil vers le bas pour lui faire reprendre de la vitesse et récupérer son assiette. Un avion risque en particulier de perdre de la vitesse lorsqu'il se cabre, ce qui est malheureusement arrivé à l'appareil du vol Rio – Paris d'Air France en 2009. Le logiciel du système MCAS est donc conçu pour faire automatiquement plonger l'avion vers le bas lorsque les données indiquent au système que l'avion est trop cabré et risque de décrocher. Ceci, sans que les pilotes concernés par l'appareil en question y soient pour quoi que ce soit, leur vigilance et leur efficacité étant présumées inférieures à celles du système automatique embarqué.

Seulement voilà. Si les données qui entrent dans le système sont mal « interprétées » par le système, le système peut « interpréter » que l'appareil se cabre indûment, alors qu'il est par exemple en phase d'ascension

normale – et indispensable – après un décollage. C'est ce qui s'est passé dans les deux cas des catastrophes des compagnies Southwest Airlines et Ethiopian Airlines à trois mois d'intervalle en 2018.

Or, cela s'est passé sans que les pilotes ne puissent rien aux opérations, pour la bonne raison qu'ils n'avaient pas été informés du fonctionnement du système dans le premier cas, et mal dans le deuxième. Outre les aspects classiquement problématiques sur le plan éthique de la question – l'information et la formation sur le désamorçage du système étaient des fonctions optionnelles payantes pour les compagnies concernées –, se pose le problème fondamental de la présupposition de l'incompétence des pilotes comparée à la compétence présumée des systèmes. Tout se passe comme si l'on présumait qu'il va tellement de soi que les systèmes électroniques embarqués sont infiniment plus « intelligents » que les humains, que l'on n'informe plus même les humains – ici, les pilotes – de ce que l'on fait faire aux systèmes et comment on le leur fait faire à la place des humains. Il y a ici non seulement problème éthique, mais problème politique. Ce n'est en

effet pas l'éthique des systèmes dits intelligents qui est en jeu. C'est celle de leurs concepteurs, bien humains – trop humains –, qui supposent savoir ce qu'est le bien d'autrui à la place d'autrui. Ce ne sont pas des machines qui sont à l'origine de la présupposition d'incompétence des humains au point de ne pas même les informer de ce que l'on met en place alors qu'ils sont directement concernés en tant qu'utilisateurs. Ce sont des femmes et des hommes dont la formation aux enjeux éthiques et politiques des systèmes qu'ils fabriquent est de toute évidence inexistante.

Nous sommes au cœur de l'éthique, dont la continuation est, selon Aristote, par d'autres moyens, la politique. Pour le dire tout à fait clairement, ce problème « éthique » que l'on attribue à ce qui s'appelle désormais l'« intelligence artificielle » se pose sur le fond d'un problème éternel de toute vie politique. C'est un problème de pouvoir, qui s'exprime par le fait que certaines et certains se croient plus « sachants » que les autres, voire sachants tout court, et qu'ils se rapportent aux autres – aux usagers – comme à des enfants incompetents. C'est le problème de toute tyrannie, de toute dictature, de toute oppression. Nous sommes exactement



dans cette même problématique lorsque l'on comprend que le produit vendu par des entreprises comme celles du GAFAM sont leurs usagers eux-mêmes, que les robots sont censés connaître mieux qu'eux-mêmes ne se connaissent. La présupposition est ici que l'identité des humains s'épuise en leurs comportements passés: l'on n'est censé « aimer » que ce que l'on a toujours déjà fait. Quel sens garde alors la notion d'« avenir » ?

La difficulté est qu'ici le problème s'exprime en douceur si l'on peut dire, au travers de l'objectivité et de la neutralité supposées des technologies et de l'intelligence censée y présider. Décidément, les problèmes éthiques que pose l'intelligence artificielle n'ont rien à voir avec les systèmes en tant que tels. Les problèmes « éthiques » liés à l'intelligence artificielle sont directement liés à l'idée que les concepteurs se font des relations entre humains et non humains. Et ici comme partout, la plus grande difficulté est que les victimes elles-mêmes de cette dynamique sont le plus souvent complices de l'oppression qui leur est imposée ou du pouvoir exercé sur elles.

S'il est urgent, pour poser correctement la question de l'éthique quand il s'agit d'intelligence artificielle, de relire le Discours de la servitude volontaire de La Boétie, il est encore plus urgent de garder à l'esprit que ce ne sont pas les machines qui sont responsables de ce que nous leur faisons faire ou de ce que nous rêvons de leur faire faire. C'est nous, les humains, qui sommes responsables des machines que nous imaginons. Les problèmes éthiques de ce qu'on appelle l'intelligence artificielle sont en fait les problèmes éthiques liés aux débordements de l'imagination humaine. Et c'est alors la République de Platon qu'il faut lire pour les poser de la meilleure façon possible. ■

*Publié initialement le 9 novembre 2020*



# ÉGALITÉ DES SEXES : L'IA EST-ELLE UNE BÉNÉDICTION OU UNE MALÉDICTION



**Stefania Santacreu-Vasut** est Professeure associée au département d'Économie à l'ESSEC Business School et au centre de recherche THEMA (CNRS). Ses travaux de recherche étudient le genre et le rôle des institutions et ont été publiés dans le *Journal of Development Economics*, *Review of Economics of the Household*, *Journal of International Business Studies*, *Journal of Economic Behavior & Organization*, entre autres. Elle est consultante pour l'OCDE, cofondatrice du projet *Gender&Finance* et le co-auteur, avec Tom Gable, du livre *The nature of goods and the goods of nature: Why anti-globalization is not the answer*.

Depuis 2010, dans le monde post #MeToo, les causes et les conséquences des inégalités de genre font l'objet d'une attention croissante de la part des universitaires, des décideurs politiques, des consommateurs et du grand public. Au cours de la dernière décennie, les préoccupations relatives à la diffusion de l'intelligence artificielle (IA) ont également attiré une attention accrue dans le débat public. L'IA est une « technologie d'usage général », dont les progrès entraînent une baisse des coûts de prédiction, notamment grâce au domaine de « machine learning » (Agrawal, Gans & Goldfarb, 2019), c'est-à-dire l'utilisation de données pour faire des prédictions. Un domaine qui sera fortement impacté par l'IA est le marché du travail, un marché où les inégalités de genre ont été particulièrement étudiées par les chercheurs en sciences sociales. L'écart salarial entre les salaires des hommes et des femmes a été déconstruit pour étudier le rôle des attributs (par exemple les différences entre les hommes et les femmes en termes d'années d'études, de choix professionnels, d'expérience professionnelle...) et le rôle de la discrimination (les différents effets des mêmes attributs). La discrimination

est souvent mesurée comme la partie de l'écart qui reste inexplicite après avoir contrôlé toutes les différences observables entre les hommes et les femmes. Une difficulté à laquelle les chercheurs sont confrontés lorsqu'ils mesurent la discrimination est de s'assurer que toutes les différences sont prises en compte, car certaines de ces différences peuvent être difficiles à mesurer et ne pas être disponibles dans les données. Comme l'IA contribue à réduire les coûts de prédiction, il n'est pas surprenant que le débat autour de l'IA ait également conduit à des questions concernant l'équité des algorithmes d'IA. Les algorithmes d'IA contribueront-ils à réduire la discrimination de genre, par exemple en améliorant les prévisions sur la productivité des travailleurs basées sur des facteurs objectifs ? Ou, au contraire, vont-ils exacerber les inégalités à l'embauche et dans la rémunération des travailleurs ? Si nous regardons au-delà du marché du travail, dans quelle mesure l'IA est-elle genrée ? Répondre à ces questions demande un effort permanent qui nécessitera une augmentation des ressources de recherche, trois considérations méritent d'être prises en compte. Premièrement, définir le point de référence correct (ou contre-factuel). Deuxièmement, la distinction



© Gettyimages - Gemlin

entre les objectifs et les prédictions des algorithmes. Troisièmement, lors de la formulation de politiques publiques, il faut tenir compte des conséquences des asymétries informationnelles entre les régulateurs et les utilisateurs d'IA.

## Le rôle du contre-factuel

Des exemples d'IA présentant des préjugés sexistes ont atteint la presse populaire, façonnant la perception du public selon laquelle l'IA conduit à des décisions discriminatoires. Pourtant, ces preuves ne suffisent pas à écarter les algorithmes de l'IA. La question clé pour les décideurs politiques n'est peut-être pas : « les algorithmes d'IA sont-ils sujets à des préjugés sexistes ? mais plutôt » Ces préjugés seraient-ils plus importants ou moindres si les algorithmes d'IA n'étaient pas utilisés ? En effet, l'alternative à l'utilisation d'algorithmes d'IA est de s'appuyer exclusivement sur le jugement humain. Comme le montrent des recherches approfondies, les décisions humaines sont souvent sujettes à des préjugés sexistes. Dans des travaux récents avec mon collègue, le professeur François Longin (Longin et Santacreu-Vasut, 2020), nous montrons que c'est le cas dans un contexte d'investissement,

un environnement où les décideurs ont pour objectif de maximiser leurs gains et ne poursuivent pas explicitement un objectif sexiste. Pourtant, les décisions d'investissement sont sujettes à des préjugés et des stéréotypes inconscients qui conduisent à des choix d'investissement biaisés, par exemple, la vente d'actions lorsqu'une femme est nommée à la tête d'une entreprise. Bien que ce ne soit pas leur objectif, les investisseurs peuvent prédire que la vente est la meilleure ligne de conduite, en partie à cause de leurs stéréotypes de genre.

## La distinction entre objectifs et prévisions

La distinction entre objectifs et prévisions est centrale dans la théorie économique. Cette distinction est extrêmement utile pour réfléchir à l'équité de l'IA (Cowgill et Tucker, 2020). Les objectifs d'un algorithme d'IA sont-ils biaisés ? Ou ses prédictions sont-elles biaisées ? Pour répondre à cette question, il est important de faire la distinction entre différents types d'algorithmes, en particulier entre ceux qui sont entièrement automatisés et ceux où un humain est « dans la boucle ».

Comme les investisseurs sur le marché financier, les programmeurs ou "l'humain dans la boucle" peuvent avoir des biais inconscients qui se traduisent par des algorithmes biaisés même lorsque l'objectif de l'algorithme n'est pas lié au genre. Les programmeurs peuvent être biaisés parce que, comme beaucoup d'entre nous, ils peuvent souffrir de biais de groupe (Tajfel, 1970), c'est-à-dire la tendance des individus à distinguer entre « nous » et « eux », profondément ancrée dans notre processus de socialisation. Comme les programmeurs peuvent être majoritairement des hommes, ils peuvent souffrir d'homophilie : la tendance à interagir avec des individus de leur propre groupe, y compris leur genre. Comment devrions-nous alors faire face à de tels préjugés ? Les outils juridiques sont-ils bénéfiques ?

## Des politiques pour contrer les objectifs et les prévisions biaisés

L'utilisation d'outils juridiques peut être utile pour lutter contre les préjugés de genre lorsque les régulateurs constatent que l'objectif d'un algorithme d'IA est biaisé. Cependant, l'utilisation d'outils juridiques stricts peut inciter les



programmeurs et les utilisateurs à créer des algorithmes moins transparents, ce qui accroît l'asymétrie informationnelle entre le régulateur et le réglementé concernant l'objectif de l'algorithme. Plus radicalement, les entreprises et les organisations peuvent décider d'éviter d'utiliser des algorithmes d'IA afin de réduire l'examen minutieux des parties prenantes et des régulateurs. Le développement d'algorithmes plus transparents peut donc conduire à un compromis entre les incitations ex ante et les incitations ex post.

Les outils politiques visant à lutter contre les biais dans les prévisions, au contraire, devront peut-être s'appuyer moins sur les outils juridiques et plus sur l'éducation. Par exemple, nous devrions former les futurs décideurs afin qu'ils déconstruisent certains de leurs propres biais et reconnaissent que les données utilisées par les algorithmes peuvent elles-mêmes contenir des biais. Pour les générations actuelles, il est important de développer des programmes de formation qui s'attaquent à la source des inégalités de genre, à savoir les préjugés humains. En résumé, la question de savoir si l'IA sera une bénédiction ou une malédiction pour lutter contre les inégalités entre les sexes dépendra de la lutte contre la racine des préjugés : pas des machines, mais des humains. ■

Publié initialement le 6 janvier 2021



#### Références

A. Agrawal, A., Gans, J., & Goldfarb, A. (2019). Economic policy for artificial intelligence. *Innovation Policy and the Economy*, 19(1), 139-159.

Cowgill, B. & Tucker, C. E. (2020). Algorithmic Fairness and Economics. Columbia Business School Research Paper, Available at SSRN: <https://ssrn.com/abstract=3361280> or <http://dx.doi.org/10.2139/ssrn.3361280>

Longin, F. & Santacreu-Vasut, E. (2019). Is gender in the pocket of investors? Identifying gender homophily towards CEOs in a lab experiment. Available at SSRN: <https://ssrn.com/abstract=3370078> or <http://dx.doi.org/10.2139/ssrn.3370078>

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223, 96-102.

# UN AGENDA DE RECHERCHE POUR ANALYSER LES DILEMMES ÉTHIQUES LIÉS AU DÉVELOPPEMENT DES SYSTÈMES DE SURVEILLANCE ET GOUVERNANCE



**Julien Malaurent** est Professeur associé au département de Systèmes d'information, sciences de la décision et statistiques à l'ESSEC Business School. Il a obtenu son doctorat en systèmes d'information de l'ESSEC Business School et de l'Université de Cergy-Pontoise en 2011. Avant d'être universitaire, Julien a travaillé pendant plusieurs années en Chine en tant que consultant ERP pour de grandes entreprises multinationales mettant en place des systèmes d'information globaux. Ses travaux de recherche, basés sur des approches qualitatives, abordent les questions liées à l'utilisation des systèmes d'information dans des contextes multiculturels, ainsi que la manière dont les organisations numérisent leurs opérations. Ses recherches ont été publiées dans les meilleures revues de la discipline telles que *Journal of Management of Information Systems (JMIS)*, *European Journal of Information Systems (EJIS)*, *Journal of Information Systems (JIS)*, *Journal of Information Technology (JIT)*. Il est également rédacteur en chef de l'*Information Systems Journal (ISJ)*.



**Thomas Huber** est Professeur assistant au département de Systèmes d'information, sciences de la décision et statistiques à l'ESSEC Business School, en France. Il a obtenu son doctorat en systèmes d'information à l'Université de Berne. Ses recherches sont principalement qualitatives et orientées vers les processus et se concentrent sur la dynamique de la gouvernance et du contrôle dans le développement inter-organisationnel de logiciels, les écosystèmes de plateformes logicielles, la collaboration numérique et la gestion des projets de transformation numérique et d'IA. Ses recherches ont été publiées dans le *Journal of Management Information Systems (JMIS)*, *Information Systems Research (ISR)*, *Information Systems Journal (ISJ)*, *European Journal of Information Systems (EJIS)*, *IEEE Software*, ainsi que dans les actes des principales conférences du domaine.

Déclenché par un événement dans votre calendrier, l'assistant personnel de votre smartphone vous conseille de partir un peu plus tôt au travail. Des actualités personnalisées s'affichent sur l'écran d'accueil de votre tablette. Votre choix de cliquer sur certaines d'entre elles ou non affinera encore plus les paramètres de votre profil. Vous arrivez au travail, répondez à vos emails, puis commandez des fournitures dont les stocks sont identifiés comme "bas" selon votre système ERP. Puis vous faites vos courses en ligne avant de rentrer chez vous. Tout au long de la journée, vous avez laissé des traces de vos données personnelles partout : les mots que vous avez prononcés à votre téléphone dès votre réveil ont été sauvegardés pour être analysés, les conditions de votre déplacement sur votre lieu de travail ont été enregistrées, la nature, le contenu et les performances de votre productivité sont aussi soigneusement tracés par les systèmes d'information de votre entreprise, la distance et les caractéristiques de votre footing ont été conservées. À cela, ajoutez le contenu de vos SMS, vos achats en ligne, vos requêtes Google, votre activité sur Facebook, etc. Vos activités privées et professionnelles sont toutes enregistrées, pour être



analysées, traitées et exploitées. Jusqu'à votre sommeil, qui sera traqué par une application de fitness dédiée.

Ceci peut vous sembler être une vision cauchemardesque d'un avenir sous surveillance - mais c'est en fait notre présent. Et cela ne fait que commencer. Cette nouvelle ère est communément rattachée au terme « capitalisme de surveillance », expression proposée par Shoshana Zuboff - professeure à l'université de Harvard. Cette explosion d'applications et services à base d'Intelligence Artificielle (IA) soulève de nouveaux dilemmes et questions éthiques au sujet de notre libre arbitre, à savoir comment positionner notre capacité d'agir, et décider face à cette surveillance ubiquitaire qui analyse nos comportements et nous propose des services et contenus sur la base de ces analyses. Dans cet article, nous tentons de présenter brièvement le statut des systèmes de surveillance basés sur l'IA et proposons quelques pistes pour analyser comment les organisations et les individus se positionnent sur ces dilemmes éthiques majeurs.

Pour commencer, il est important de rappeler que les systèmes à base d'IA, tels que les services offerts par Google ou

Facebook, ont pour modèle économique de recueillir, traiter et commercialiser les informations collectées sur nos activités et habitudes de vie. Cela implique des données telles que nos comportements sur les réseaux sociaux, nos habitudes d'achat et transactions effectuées par carte de crédit. Mais la prochaine génération de ces systèmes ira bien au-delà. À mesure que les technologies émergentes telles que les capteurs contextuels vont se développer, les caméras de reconnaissance faciale et les systèmes d'analyse de foule deviendront plus performantes, et ces acteurs disposeront alors d'un éventail plus large de données relatives à nos actions, non seulement dans le monde numérique, mais aussi dans le monde physique. Ces données atteindront un niveau de granularité sans précédent et pourront être utilisées non seulement pour prédire le comportement individuel de chacun mais aussi pour le façonner, le produire, le modifier, voire le contrôler.

Ce n'est pas seulement le secteur privé, aujourd'hui porté par les géants technologiques américains (GAFAM) et chinois (BATX), mais aussi le secteur public qui accélère le développement et l'usage de ces systèmes d'information à base d'IA. De nombreux pays ont

d'ores et déjà commencé à utiliser des systèmes de gouvernance à grande échelle, alimentés par l'IA. Le système du Crédit Social Chinois en est un parfait exemple. Ce système en cours d'expérimentation est fondé sur une surveillance ubiquitaire de sa population afin d'enregistrer et analyser toutes les caractéristiques des ménages, couvrant la consommation d'énergie, la scolarité des enfants, et le sommeil de chacun. En augmentant la surveillance des comportements humains et en quantifiant les actes les plus personnels, les nations qui développent ces systèmes ont pour volonté d'étendre leurs outils de gouvernance.

## Deux dilemmes éthiques des systèmes à grande échelle alimentés par l'IA

Malgré les nombreuses promesses de ces systèmes de surveillance et de gouvernance alimentés par l'IA, ils soulèvent également d'importantes questions éthiques, non seulement du point de vue des citoyens et de la société au sens large, mais aussi du point de vue des concepteurs et des





développeurs. Le premier dilemme concerne la collecte et l'exploitation des données. D'une part, les développeurs de ces systèmes à base d'IA sont fortement incités à collecter et à exploiter des quantités de données toujours plus importantes. La raison de cette insatiable soif de données est la logique technique des systèmes d'IA modernes. Plus ces systèmes sont alimentés en données, plus ils deviennent précis dans leurs analyses, comportements ou recommandations. D'autre part, plus les fournisseurs de systèmes d'IA utilisent ces données pour façonner et contrôler nos comportements, plus ils restreignent notre libre arbitre. Nos capacités de jugement et d'action sont alors largement affectées par des algorithmes et processus de classification (Zuboff, 2019). Par conséquent, un important défi éthique consiste à trouver le juste équilibre entre les besoins de gouvernance que ces systèmes sont censés servir et le droit à conserver notre libre arbitre.

Un deuxième défi éthique concerne celui de la responsabilité des recommandations et actions fournis par les systèmes à base d'IA. Si un tel système suggère un comportement indésirable ou discriminatoire, qui en est responsable ? Nous devons arrêter de penser que les machines sont mauvaises ou maléfiques par défaut. Les algorithmes ne font pas la distinction entre les « bonnes » et les « mauvaises » personnes ou entre les bons et les mauvais comportements par eux-mêmes. Ces systèmes ont été produits par des professionnels de l'informatique, c'est-à-dire par des concepteurs, des développeurs et des ingénieurs. Ces professionnels les ont alimentés avec des règles de classification et de regroupement ou du moins avec des données qu'ils ont jugé appropriés pour construire ces règles de manière inductive. Pourtant, il n'est pas évident qu'il faille attribuer la responsabilité aux seuls ingénieurs. Les limites de la responsabilité morale sont floues

par le fait qu'en raison de leur nature intrinsèque, les systèmes artificiellement intelligents sont dans une certaine mesure autonomes ; ils suggèrent des comportements et actions qui vont au-delà de leur programmation initiale.

### Examiner la gestion des dilemmes éthiques à l'ère de l'IA

L'ESSEC a récemment structuré sa stratégie (RISE) autour de trois piliers stratégiques : Enlightening Entrepreneurship, Together, et le Metalab. Le premier vise à décupler la puissance de notre écosystème lié à l'entrepreneuriat et l'innovation. Le second consiste en la création de connaissances et savoirs qui aideront les générations actuelles et futures à agir non seulement avec un sens des affaires, mais aussi avec une responsabilité sociale et morale (Together). Le troisième consiste en la création de programmes et débats

qui aideront les dirigeants à naviguer dans la « tempête » des changements induits par les technologies numériques (Metalab). À travers des enseignements et participations aux débats publics sur le développement et l'usage de l'IA, le Metalab souhaite influencer sur la prochaine génération de ces systèmes pour que ceux-ci ne profitent pas seulement à quelques conglomérats privés, ainsi qu'à prévenir toute forme de système qui deviendrait rapidement hors de contrôle.

C'est dans ce contexte que nous avons récemment initié un projet de recherche qui a pour objectif l'examen détaillé de l'une des plus grandes initiatives d'IA au monde, à savoir le système de crédit social chinois. Par cette étude aussi bien sociétale que technologique, nous nous efforçons d'identifier et analyser les dilemmes éthiques auxquels nous allons être confrontés en tant qu'utilisateurs mais aussi concepteurs et/ou ambassadeurs de ces technologies.

À travers cette étude à grande échelle, nous espérons en outre faire progresser considérablement notre compréhension sur la manière dont les aspects techniques, managériaux et éthiques s'entremêlent et entraînent ensemble des conséquences tant voulues qu'involontaires. Pour finir, nous espérons que ce projet de recherche produira des leviers managériaux actionnables pour concevoir, gouverner et utiliser ces systèmes dans le respect de règles éthiques et morales qui doivent, en parallèle, être redéfinies dans les contours d'une société hyper-technologique. ■

Publié initialement le 4 janvier 2021

#### Note

Cet article présente les principaux aspects de motivation d'un grand projet de recherche mené par les professeurs Huber et Malaurent. Cette recherche est financée par la Fondation ESSEC dans le cadre de la bourse White Project obtenue en janvier 2020. Pour plus de détails, veuillez contacter les auteurs aux adresses suivantes : huber@essec.edu et malaurent@essec.edu

#### Référence

Zuboff, S. (2019) *The age of surveillance capitalism: the fight for the future at the new frontier of power*, Profile Books.



# VERS UNE POLITIQUE DE SÉCURITÉ DES ALGORITHMES



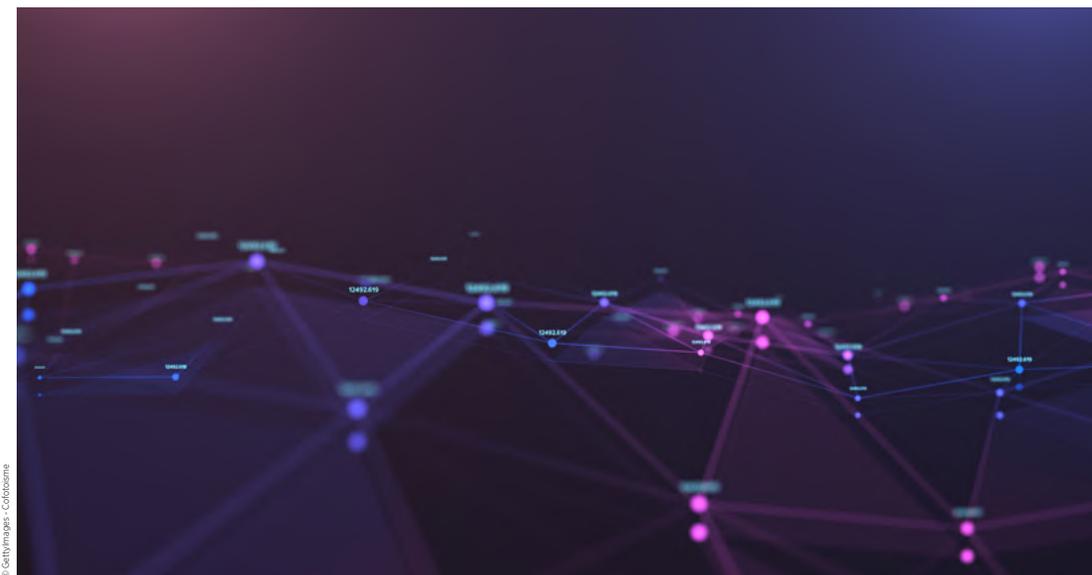
**Guillaume Cheillon** est Professeur au département des Systèmes d'information, sciences de la décision et statistiques à l'ESSEC Business School. Il dirige le Master ESSEC/CentraleSupélec Data Sciences & Business Analytics. Son domaine de recherche est la théorie économétrique avec des applications à la prévision et à l'apprentissage statistique en économie et en finance. Ses intérêts concernent l'étude des aspects dynamiques dans les réseaux à grande échelle, la dépendance dans les environnements instables et l'apprentissage par les agents économiques. Il a travaillé sur des applications en économie, finance, énergie, sur les réseaux sociaux, le réchauffement climatique et la dynamique des pandémies. Guillaume a obtenu un MPhil & un DPhil en Économie à l'Université d'Oxford, & un Diplôme d'Ingénieur à l'École des Mines de Paris. Il fait partie du réseau d'experts de l'OCDE sur l'IA et a été chercheur ou professeur invité à Brown, Oxford, NYU, Keio, UNSW Sydney & à la Fed de NY.

Alors qu'Elizabeth Warren et Bernie Sanders, tous deux sénateurs et candidats aux primaires présidentielles du Parti Démocrate américain, ont plaidé pour le démantèlement des entreprises « Big Tech » (les géants américains du web)<sup>1</sup>, la Commissaire Européenne à la Concurrence Margrethe Vestager a déclaré à plusieurs reprises au cours des dernières années que, si elle partageait les mêmes objectifs de protection et de liberté des utilisateurs, la solution du démantèlement par les lois antitrust ne lui semblait pas efficace<sup>2</sup>. Elle estime plus utile de combiner la promotion de la concurrence avec des contraintes réglementaires comme, par exemple, le règlement général sur la protection des données (RGPD) ou les Directives récentes sur les marchés et services numériques. Mark Zuckerberg a lui-même parfois appelé à une plus grande régulation des réseaux

sociaux, tandis que son cofondateur de Facebook, Chris Hughes, va plus loin et préconise de démanteler l'entreprise<sup>3</sup>. Que faut-il donc faire ?

Pour améliorer le contrôle par les pouvoirs publics, certains, comme Hannah Fry, une mathématicienne de l'University College of London qui a publié en 2018 un livre à succès sur les données et les algorithmes<sup>4</sup>, proposent d'établir une autorité de régulation des algorithmes, inspirée par la Food and Drug Administration américaine. Une telle proposition mérite attention, car elle suggère que les algorithmes de recommandation – aujourd'hui principalement ceux des réseaux sociaux, mais demain tous ceux regroupés sous le terme d'Intelligence artificielle – présentent un potentiel de danger qui doit être évalué avant leur commercialisation.

1 - <https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c>  
2 - [https://eeas.europa.eu/delegations/united-states-america/43309/commissioner-margrethe-vestager-press-conference-washington-dc\\_en](https://eeas.europa.eu/delegations/united-states-america/43309/commissioner-margrethe-vestager-press-conference-washington-dc_en)  
<https://www.competitionpolicyinternational.com/eu-vestager-says-breaking-up-facebook-would-be-a-last-resort/>  
3 - <https://www.ft.com/content/0af70c80-5333-11e9-91f9-b6515a54c5b1>  
<https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html>  
4 - Hannah Fry (2018), Hello World: How to be human in the age of the machine, 2018, London: Penguin.



© Gettyimages - Colobome

Selon cette analogie sanitaire, les algorithmes apportent un bénéfice, mais peuvent avoir des effets secondaires néfastes (enfermement dans des bulles informationnelles, comportements de dépendance, effondrement des pratiques démocratiques...). C'est également le cœur de la critique formulée par le documentaire sponsorisé par Netflix, Derrières nos écrans de fumée (The Social Dilemma), qui appelle à un changement radical des modèles économiques afin de s'éloigner du modèle actuel d'« extraction de l'attention humaine ». Ce documentaire fait référence au Center for Humane Technology, dont le président Tristan Harris a écrit dans le Financial Times (en mars 2020)<sup>5</sup> que la question clé n'est pas seulement celle de la propriété et de la revente des données, mais bien du fonctionnement des algorithmes des plateformes de réseaux sociaux dont l'objectif est de maximiser l'engagement personnel à tout prix. Il appelle à la réglementation de ces plateformes en tant que « services publics de l'attention humaine », soumis

à une licence qui garantirait qu'ils fonctionnent dans l'intérêt public.

## Attention humaine et bulles sociales

Selon cette proposition, une agence indépendante pourrait analyser les algorithmes ex ante via une « évaluation de l'impact social » et, le cas échéant, permettre leur déploiement. Hannah Fry et Tristan Harris semblent aller au-delà des propositions de Bernie Sanders, d'Elizabeth Warren et de Margrethe Vestager d'une surveillance par l'autorité publique : ils demandent un contrôle administratif a priori.

La justification de ce contrôle est claire : les recommandations algorithmiques sont là pour nous faire réagir rapidement, non pour nous présenter toutes les alternatives pertinentes qui nous permettraient de prendre une décision éclairée. En effet, les plateformes sociales telles que Twitter et Facebook sont financées par la publicité et

leur intérêt est donc de maximiser l'intensité de leur réseau, l'audience des interventions et l'amplitude des réactions afin de recueillir toujours davantage d'informations sur les goûts, les intérêts et les préférences de leurs utilisateurs. Cela leur permet de nous suggérer les publicités les plus pertinentes ou, comme Netflix, qui modifie les couvertures des films et des séries en fonction de nos réactions, de nous présenter les suggestions qui auront le plus d'influence sur nos comportements

Dans le contexte des réseaux sociaux, cela génère le phénomène du type « un homme a mordu un chien » selon lequel l'information la plus référencée, et donc partagée, n'est pas nécessairement la plus pertinente, mais la plus surprenante<sup>6</sup> (en inversant, ici, le plus habituel « un chien a mordu un homme »). De nombreux médias en ligne se font ainsi connaître à travers une course aux « nouvelles » provocantes ou surprenantes (qui n'a pas vu un titre accrocheur « vous ne croirez pas ce qui est arrivé à... ») qui

5 - <https://www.ft.com/content/abd80d98-595e-11ea-abe5-8e03987b7b20>  
6 - Pour une application Macroeconomics, voir Nimark (2014), <https://www.aeaweb.org/articles?id=10.1257/aer.104.8.2320>



sont finalement peu informatives. En retour, les plateformes sociales ne recueillent pas d'informations sur nos centres d'intérêts réels. Le message de Derrières nos écrans de fumée est que la maximisation de l'engagement ne devrait pas être mesurée par le temps passé en ligne et le nombre d'interactions. La mesure devrait plutôt contenir une évaluation de la qualité de ces interactions.

Faute d'une telle focalisation sur la qualité, le succès des recommandations est actuellement mesuré par l'intensité de l'engagement, à savoir notre réponse aux stimuli. C'est pourquoi les algorithmes tâchent de modéliser nos préférences et intérêts en s'appuyant, pour mieux nous comprendre, sur notre historique de comportement, nos lectures et actions en ligne. Il s'agit ensuite de prédire nos futures réactions. Ces prédictions sont toutefois entachées d'erreurs, car elles reposent sur les informations partielles que nous fournissons par nos parcours observés, non sur l'éventail plus large de nos intérêts potentiels. Ce faisant, elles réduisent la diversité des suggestions et notre exposition aux idées qui nous dérangent : agissant comme des mécanismes de renforcement, elles peuvent nous enfermer dans une bulle informationnelle, une chambre d'écho. C'est la principale critique formulée par Tristan Harris et le Center for Humane Technology : les bulles induites par les plateformes de réseaux sociaux peuvent nous éloigner les uns des autres, favoriser les divisions et abîmer le tissu social. Derrière nos écrans de fumée, présage par conséquent la fin de la démocratie et utilise l'exemple des Gilets jaunes qui se sont organisés et ont partagé leurs informations sur Facebook et WhatsApp. S'il est vrai que les fausses nouvelles qui résident dans ces bulles sociales rendent le dialogue difficile, les historiens des mouvements sociaux soutiendraient toutefois sans doute que les groupes progressistes développent généralement leurs propres récits et que ceux-ci diffèrent

de ceux des médias établis (on pourrait prendre les exemples aux États-Unis de la lutte pour les droits civils dans les années 1950-1960, ou ceux des minorités sexuelles dans les années 1970-1980). Les arguments fondés sur le mouvement des Gilets jaunes sont donc discutables : la question qui importe est celle de l'intensité et de la prévalence généralisée de ces bulles, plutôt que de leur simple existence.

## Des agences nationales de sécurité des algorithmes ?

La création d'une Agence nationale — ou européenne, ou internationale — de sécurité algorithmique, aussi intéressante qu'elle puisse paraître au premier abord, oublie un élément essentiel : les algorithmes des plateformes sociales ne sont pas seulement des objets scientifiques (mathématiques ou informatiques) qui peuvent être évalués a priori, mais ils doivent être analysés dans un contexte de sciences sociales à travers leurs impacts à moyen ou long terme. Sans toutefois présenter les conséquences dramatiques de Skynet dans la série de films Terminator, tout algorithme social — in fine une simple recette pour obtenir un certain résultat (un comportement) en utilisant certains ingrédients (stimuli) — échappe automatiquement au contrôle de son concepteur. La question n'est donc pas uniquement de savoir si les plateformes sociales sont indifférentes à l'étude des conséquences de leurs outils, comme semblent le laisser entendre les participants à Derrières nos écrans de fumée.

En effet, les algorithmes de réseaux sociaux ne sont qu'une nouvelle méthode pour atteindre un objectif ancien ; c'est, de fait, le principe même de toute politique publique : influencer le comportement des individus. L'histoire est parsemée de nos échecs à cet égard. Depuis la création d'instituts

de statistiques, comme le Bureau de recensement aux États-Unis ou l'INSEE en France et le développement des techniques de sondage, les pouvoirs publics et les entreprises privées utilisent les données et les statistiques pour analyser les comportements des citoyens et des consommateurs : ils tentent également de modifier ces comportements pour obtenir des résultats précis (augmentation de la croissance économique, réduction de la pauvreté, amélioration des ventes...).

Lorsqu'il s'agit d'influencer un être humain, la difficulté réside toutefois dans le fait que le comportement de celui-ci évolue en fonction des influences qu'il reçoit. Nous humains sommes comme des machines qui changeraient d'usage, de forme ou de mode de fonctionnement, dès qu'on essaie de nous actionner.

## Prendre en compte les réactions humaines

Les sciences sociales et économiques étudient depuis longtemps les influences réciproques entre les individus et leur environnement et, dans ce contexte, la question de la modélisation et du contrôle. Ici, la question importante n'est pas de savoir si un algorithme est néfaste : en tant que recette, il est conçu dans un but précis et fonctionne généralement assez bien à court terme. Toutefois, il s'agit d'une recette partielle qui n'utilise qu'une fraction (même si elle est importante) des ingrédients possibles. Lorsqu'il s'agit de susciter l'engagement sur les réseaux sociaux, comme les notions de véracité ou de qualité sont absentes des algorithmes actuels et que seule la popularité est prise en compte, la désinformation se développe. Ainsi, l'algorithme atteint son objectif à court terme, mais son impact à moyen terme (polarisation de l'information, manque de contradiction et de hiérarchisation) n'est pas pris en compte. Il s'agit ici de la notion économique d'externalité, selon laquelle les entreprises internalisent

les bénéfices de la technologie, mais externalisent les conséquences négatives (la pollution par exemple, dans un contexte industriel) à la société dans son ensemble<sup>7</sup>.

En sciences sociales, on sait que tout stimulus modifie non seulement l'individu affecté mais aussi le contexte dans lequel celui-ci opère. Imaginons que vous analysiez les comportements d'un groupe d'individus et que vous les modélisiez à l'aide d'un algorithme : lorsque vous essaieriez d'utiliser cet algorithme pour influencer ces personnes, vous modifieriez de fait leur environnement puisque quelqu'un — en l'espèce, vous — essaie d'affecter leur comportement « habituel ». Cette modification contextuelle génère à son tour de nouvelles réactions qui peuvent potentiellement rendre l'algorithme inutile ou même contre-productif<sup>8</sup>. Par exemple, une célèbre controverse récente sur Internet concernait l'utilisation par les compagnies de recherches en vue d'identifier les vacances que les internautes prévoient, et ce dans le but d'augmenter le prix des vols proposés. Lorsque ceci a été connu, de nombreux internautes ont joué avec leurs recherches de vols afin de perturber les cookies et d'obtenir, contrairement aux prévisions algorithmiques, des prix plus bas<sup>9</sup>.

Aussi la dimension humaine n'est-elle pas suffisamment prise en compte dans les algorithmes issus des sciences de l'ingénieur où les individus sont considérés comme des boîtes noires : leurs pensées ne sont pas perçues, mais leurs conséquences sont mesurées à

travers les actions qui en résultent. En réalité, les humains réfléchissent aux influences qui s'exercent sur eux, et ils peuvent les contrecarrer.

## Garantir la durabilité à long terme

Dans un tel contexte, il semble illusoire pour une autorité administrative de contrôler ex ante les outils de l'intelligence artificielle, car leurs conséquences à moyen terme sont presque imprévisibles compte tenu du nombre d'acteurs exerçant une influence. Une réponse adéquate à la question posée par Derrière nos écrans de fumée peut en revanche être trouvée dans un contexte éloigné de la santé et des médicaments, celui de la maîtrise de l'inflation.

En effet, les pouvoirs publics ont longtemps cherché à éviter le double écueil d'une inflation trop élevée (hyperinflation entraînant l'instabilité politique dans les années 1920) ou trop faible (déflation entraînant l'appauvrissement vécu au cours des années 1930). Une inflation modérée est optimale, mais c'est un équilibre instable, résultat des décisions individuelles de millions d'individus et d'entreprises, décisions qui sont elles-mêmes le résultat de la perception qu'ont les gens de leur environnement (passé, présent et futur) et des décisions des autres (concurrents, fournisseurs, clients...).

Après avoir pensé que les gouvernements pouvaient contrôler directement les prix (par exemple, le prix du pain en France jusque dans les

années 1980) ou les outils de politique, un consensus s'est forgé dans les milieux universitaires au cours des cinquante dernières années qui consiste à penser que l'agence chargée de la lutte contre l'inflation, la Banque centrale, se doit d'être indépendante du gouvernement et en pleine possession des outils pertinents — non pas ceux d'un contrôle direct des prix individuels des biens et services, mais ceux qui influencent la prise de décision individuelle (via les taux d'intérêt) ou qui permettent la supervision des principaux opérateurs de marché (banques et institutions financières). Dans ce contexte, le rôle du gouvernement consiste à fixer les objectifs de la politique monétaire (une inflation modérée et, aux États-Unis par exemple, le plein emploi). Les banques centrales ont été rendues indépendantes pour convaincre la population de leur capacité à poursuivre les objectifs à moyen terme qui leur ont été prescrits et de s'émanciper des considérations politiques de court terme — celles qui peuvent favoriser la réélection d'un gouvernement au détriment de l'intérêt de la nation. Nous avons par exemple assisté à une tentative de changement de ce statu quo lorsque le président Trump a récemment menacé de remplacer le président de la Réserve fédérale afin de le forcer à modifier sa politique de taux d'intérêt<sup>10</sup>.

7 - Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *American Economic Review*, 75(3), 424-440.  
Wattal, S., Racherla, P., & Mandviwalla, M. (2010). Network externalities and technology use: a quantitative analysis of intraorganizational blogs. *Journal of Management Information Systems*, 27(1), 145-174.  
Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211-36.  
8 - C'est un exemple de la célèbre critique de Lucas qui a été introduite en économie par Robert E. Lucas (Prix Nobel d'économie 1995) : Lucas, R. E. (1976). *Econometric policy evaluation: A critique*. In *Carnegie-Rochester Conference Series on Public Policy* 1(1), 19-46.  
9 - <https://time.com/4899508/flight-search-history-price/>  
<http://www.businessinsider.fr/us/clear-cooking-when-searching-for-flights-online-2015-9>  
10 - <https://www.newyorker.com/news/our-columnists/the-high-stakes-battle-between-donald-trump-and-the-fed>

## Une banque centrale des algorithmes

Plutôt que d'instaurer un système d'autorisation administrative de mise sur le marché, i.e. un contrôle a priori par une Agence de sécurité des algorithmes, les technologies de l'intelligence artificielle pourraient être mieux maîtrisées par une autorité indépendante qui supervise directement les sociétés d'IA et impose

certains algorithmes de base. Cette autorité pourrait, par exemple, contrôler des algorithmes « essentiels », posséder la capacité de les modifier et obtenir des mesures d'impact quotidiennes (comme la Banque centrale vérifie chaque nuit que les banques privées équilibrent leurs comptes). Cette agence indépendante, cette Banque centrale d'algorithmes, pourrait ainsi introduire un objectif de moyen terme et permettre l'évolution de la société

en accord avec les objectifs fixés par le gouvernement. Elle pourrait également surveiller le degré de concentration du secteur afin d'éviter l'émergence d'entreprises "trop grandes pour faire faillite" (Too big to fail) qui mettent en péril l'ensemble du système.

La capacité de cette entité à agir directement, son indépendance et sa focalisation sur des objectifs explicites contribueraient à favoriser la confiance

systemique de tous les agents, citoyens comme entreprises. La confiance est le facteur clé qui permet de mieux anticiper les réactions des individus aux stimuli reçus : elle améliore la réactivité et facilite la résolution des problèmes posés par les bulles d'information et la désinformation. Comme pour l'innovation financière que la réglementation limite afin de réduire le risque de crise économique majeure (et qui survient néanmoins

lorsque la régulation est réduite), le développement de l'intelligence artificielle pourrait ainsi être légèrement ralenti, en contrepartie du maintien d'un objectif d'intérêt public et du bénéfice de soutenabilité à long terme. ■

*Publié initialement le 21 décembre 2020*



# TELLEMENT DE MANIÈRES DE SE TROMPER SUR LE FUTUR DE L'IA, MAIS ELLE NOUS DÉPASSERA



**Fabrice Cavarretta** est Professeur associé de leadership et d'entrepreneuriat à l'ESSEC. Coordinateur du séminaire doctoral en Entrepreneuriat, il a développé plusieurs programmes d'intrapreneuriat en formation continue. Ses recherches académiques portent sur la logique utilisée par les managers pour développer de nouvelles entreprises, et sur l'application de l'intelligence artificielle à l'analyse des individus et organisations. Le professeur Cavarretta a publié le livre *Oui! La France est un paradis pour entrepreneurs* (Plon, 2016) dans lequel il aborde la spécificité des écosystèmes entrepreneuriaux nationaux.

**E**n 2015, je décidais d'explorer les nouvelles capacités de l'intelligence artificielle dans le management. L'IA venait de redevenir populaire, il était temps de comprendre si — et comment — les outils avaient évolué. L'IA pourrait-elle être utilisée au-delà de récents succès dans des domaines très spécifiques tels que la reconnaissance d'images ? Commencerait-elle à avoir assez de souplesse pour découvrir de complexes relations dans des données floues, par exemple en entreprise entre les individus, les équipes, la performance, les émotions, etc. ?

Après quelques années d'exploration, j'ai décidé de reporter ce projet sine die et de me préparer à un nouvel hiver de l'intelligence artificielle, une de ces phases de lent progrès, où l'on se met à douter sur le décollage ultime de la technologie.

Bien entendu, les algorithmes actuels d'IA impressionnent. Leurs capacités à répondre à des questions prédéfinies sont extraordinaires : par exemple, ces algorithmes savent maintenant très bien identifier les chats, les individus, les voitures, dans des situations toujours plus naturelles.

Mais pour chaque annonce de nouvelle application de l'IA, on compte encore plus de cas où nombre d'obstacles empêchent sa mise en œuvre pragmatique : données non accessibles, non formatées ou trop bruyantes ; manque de capacité analytique et de personnel formé ; technologie pas adaptée ; problèmes juridiques ; préoccupations éthiques, etc. La liste des bonnes raisons pour ne pas utiliser l'IA — même dans un domaine étroit — pourrait devenir un sujet de recherche en soi.

Paradoxalement, malgré tous ces obstacles actuels, l'IA atteindra probablement le Graal, une forme d'intelligence indistinguishable de celle des humains.

## Une IA pourrait-elle faire preuve d'intelligence au sens humain ?

Les obstacles pratiques actuels sont faibles à côté de l'immense difficulté à reproduire des processus cognitifs typiquement humains. Par exemple, ni l'intuition -- des cognitions émergentes du foisonnement historique d'un individu - ni les émotions - qui permettant l'autonomie



par les « cognitions motivées » - ne sont actuellement reproduites dans les systèmes actuels d'IA.

Si l'IA peut répondre à des questions précises dans des domaines précis avec des données bien identifiées, qu'en est-il de la capacité à trouver par soi-même des relations utiles dans des ensembles complexes de données hétérogènes ? C'est bien par cette capacité à faire émerger une pensée de manière floue et motivée que l'humain se distingue encore clairement des machines actuelles.

Pour le futur, les experts se déchirent : certains vous parlent de l'IA comme d'un fait accompli qui va envahir le monde très bientôt et surpassera l'humain rapidement. D'autres jurent que jamais une machine ne fera preuve d'une intelligence comparable à celle des humains. Qui croire alors ? Notre monde surestime-t-il ou sous-estime-t-il la vitesse à laquelle l'IA va progresser vers l'intelligence au général ?

## Une IA très étroite dans le court terme, très large après des progrès à venir...

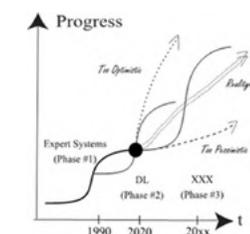
Paradoxalement, la bonne réponse est... « les deux ». Ce paradoxe dérive d'un phénomène classique pour toute innovation. Quand un domaine évolue, le changement n'a pas lieu de manière linéaire, progressive. La réponse d'une communauté humaine, technique, sociale à une évolution suit ce qu'on appelle une courbe en S : au début, le progrès est très lent, puis il y a une explosion de l'usage, et ensuite cela s'apaise à nouveau. Dans chaque contexte, de telles évolutions s'accumulent par vagues au fil du temps, elles se combinent d'une manière parfois difficile à interpréter. Visualisons cela dans la figure attachée, où trois courbes en S représentant trois vagues d'innovations qui s'enchaînent.

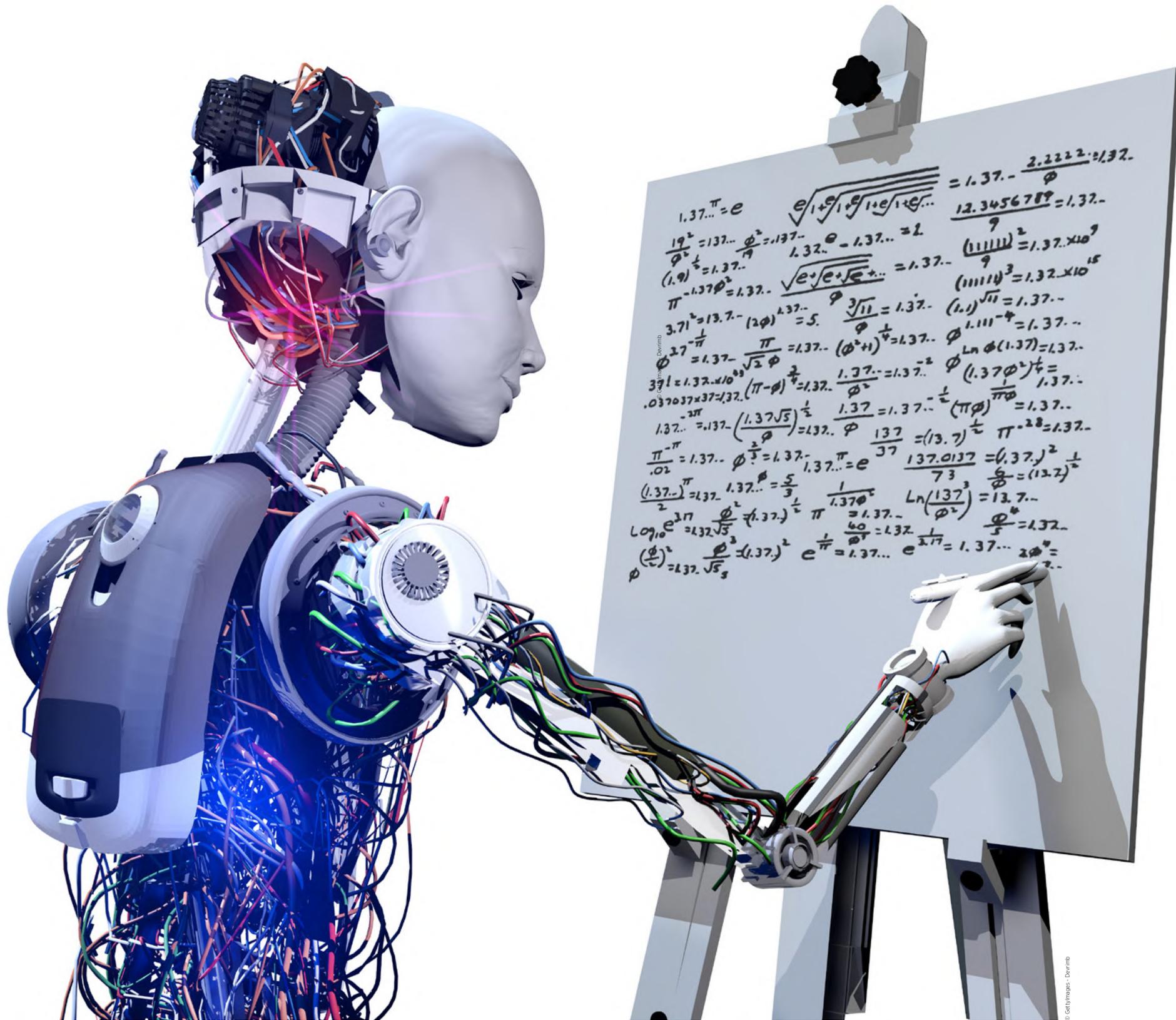
À tout moment, plusieurs prévisions sont possibles selon ce qu'on considère. Si on observe la première vague d'innovation (n°1) qui ne progresse plus, on va anticiper que la technologie ne mènera pas à grand-chose. Ceci constitue une projection trop pessimiste, et on

fait paradoxalement de même si on observe l'innovation naissante (vague n°3), donc qui n'avance pas encore significativement.

Par contre, si on projette vers le haut la vitesse de l'innovation en cours (vague n°2), alors on va surestimer le progrès dans une projection trop optimiste.

Tous ces raisonnements sont faux : en effet, le progrès dû à l'innovation n°2 va se calmer, mais par contre le progrès attendu de l'innovation n°3 aura bien lieu, et sera probablement suivi par les progrès d'autres innovations à venir n°4, n°5, n°6, etc. C'est une succession d'innovations, y compris toutes celles encore à venir et non planifiables qui constituent la trajectoire réelle. Paradoxalement, elle se situe quelque part « au milieu », dans une trajectoire toujours difficile à imaginer, car ne collant à aucune des tendances du moment !





## L'IA est au milieu du gué

Imaginer un futur de l'IA est donc difficile, en tension entre des perceptions contradictoires... Afin de prendre du recul, souvenons-nous que l'IA appartient au domaine de l'informatique, dont la conception s'est déroulée sur plusieurs siècles.

L'idée de calculateur automatique remonte au siècle des Lumières, mais il était alors imaginé que cela se baserait sur des machines mécaniques — forcément très limitées. Au milieu du XX<sup>e</sup> siècle, l'électromécanique puis la première génération d'électronique fait changer l'informatique d'échelle, la nouvelle capacité des calculateurs rend possibles des applications pratiques en entreprise (émergence d'IBM par exemple). À la fin du XX<sup>e</sup> siècle, la miniaturisation des semi-conducteurs donne lieu à une nouvelle vague de progrès et d'adoption massive (émergence de l'informatique de masse, avec le PC par exemple). Enfin, depuis le début du XXI<sup>e</sup> siècle, la diffusion généralisée de machines et de réseaux a encore changé la donne, avec la vague internet, la mobilité, les objets connectés.

À chaque étape, l'informatique a suscité autant d'occasions de faire preuve de pessimisme (« mais la machine est très limitée! ») que d'optimisme (« ces machines vont envahir le monde »). Alors, où en est-on pour l'IA? très précisément au milieu du gué, dans la position pleine d'incertitude de n'avoir observé que deux vagues de progrès.

Une première vague initiée — dans les années 50 et en plein essor dans les années 80-90 — a supposé que les machines devraient faire preuve d'intelligence de manière dite « symbolique ». L'IA était modélisée comme le travail d'un ingénieur qui manipule des concepts par des logiques discrètes. Par exemple, dans des systèmes experts contrôlant des centrales nucléaires, l'IA prenait la forme de système de règles de la forme : « si (alarme = déclenchée & vanne = ouverte) alors (fermer vanne) ». Ceci était fort excitant, donna lieu à beaucoup d'études et quelques implémentations réussies.



Malheureusement, cela mena à une impasse, car cela nécessitait que l'humain l'encode entièrement et explicitement, une forme circulaire d'intelligence qui ne fit pas illusion très longtemps. Le tournant des années 2000 fut un premier hiver pour les spécialistes de l'IA, une phase où on l'on rêvait plus de grands progrès à venir.

Au début des années 2010, la donne est bouleversée : des résultats impressionnants sont obtenus par des chercheurs travaillant depuis des décennies sur des modèles statistiques appliqués principalement à la vision en robotique. Paradoxalement, ceci n'avait pas été considéré de l'IA, car à l'époque celle-ci n'était envisagée que sous la forme de manipulations symboliques. Leur méthode prend le nom de « deep learning » (DL), car elle est basée sur un grand nombre de neurones virtuels. Avec le DL, les machines peuvent exécuter des tâches relativement sophistiquées, par exemple identifier un chat dans les millions de pixels d'une image.

Mais surtout, ces techniques permettent un apprentissage relativement automatiquement — qu'on appelle alors « Machine Learning » (ML) — à condition qu'on puisse fournir à la machine des ensembles de données massives et que l'humain désigne ce qu'il faut chercher. Par exemple, avec un grand ensemble de photos qui ont été taguées pour savoir si elles contiennent un chat ou pas, on peut entraîner automatiquement un algorithme à « repérer des chats » dans des photos en général.

Nous sommes en plein dans le boom de cette deuxième génération d'IA, et les nombreux résultats sont impressionnants. Si on se focalise seulement sur ces succès récents du ML, on pourrait en conclure qu'on va aller au ciel, que l'IA va envahir la planète ... maintenant ! Ceci est une illusion, c'est la « projection trop optimiste » dans la figure, qui surestime le futur.

En effet, la force du ML vient de sa grande simplicité, d'où une critique que cette IA revienne à n'être que de la statistique sophistiquée. Cette expression est péjorative, car elle rappelle que la statistique n'est qu'une méthode par laquelle une hypothèse faite par un humain peut être validée sur des données. Le ML est effectivement principalement de la statistique, car l'humain est toujours censé jouer le rôle central dans l'imagination des relations entre les entités du monde, et la machine vérifie automatiquement, ou à la rigueur fait le tri automatiquement. Remarquez bien que la machine n'a encore rien « imaginé » dans cette approche...

Le ML actuel se cantonne donc à des tâches prédéfinies, on ne peut lui déléguer des processus complexes comme nettoyer et choisir les données, détecter différentes relations entre variables, identifier les variables qui pourraient être intéressantes. Et surtout, cette IA ne peut suggérer des mécanismes explicatifs : par exemple déterminer que « la photo contient un chat parce que c'est un animal avec oreille pointue et moustache ». Non seulement cette technologie ne suggère rien, mais en plus, quand sa statistique s'appuie sur un ou des éléments clairs, elle n'en est pas capable de nous le dire, car les algorithmes actuels ne sont pas encore conçus pour cette explicitation.

Pour toutes ces questions, le faible avancement des technologies actuelles suggère qu'on n'y arrivera jamais, qu'une intelligence comparable à celle de l'humain sera trop compliquée, qu'on espère trop de la machine. Cette perception est tout autant une illusion, c'est la « projection pessimiste » qui sous-estime le futur.

La trajectoire réelle, comme toute prédiction en situation d'incertitude, est difficile à cerner, quelque part entre ces deux tendances. L'IA devra accomplir une succession de progrès pour arriver un jour à une vraie forme d'intelligence. Ces progrès auront très probablement

lieu, mais entre temps, nous passerons probablement par d'autres « hivers de l'IA », ces périodes où les choses n'avancent plus trop, où plus personne ne croit que les lendemains chantent.

## Trajectoire vers une intelligence artificielle généralisée

Pour l'instant, les technologies de ML sont monocouches, l'humain spécifie les entrées et utilise les résultats. Parfois, un ingénieur peut décider manuellement de prendre un résultat d'un premier algorithme et d'en faire l'entrée d'un deuxième algorithme. Mais dans le futur, il faudra mettre ce bouclage à l'intérieur d'un processus de ML, que ces relations s'enchaînent en cascade comme le fait le cerveau. Mais pour cela, voici les progrès principaux à attendre.

Le premier progrès serait la prise en compte du symbolique dans le cadre des techniques d'apprentissage automatiques actuelles. Le ML est efficace pour l'instant à traiter beaucoup de données plutôt continues (des sons, des images) afin d'en deviner une structure (distinguer les images avec chat de celles sans chat). Malheureusement, pour l'instant, le ML n'est pas adapté à l'analyse des données symboliques, c'est-à-dire de données discontinues, et encore moins quand elles sont en petit volume.

Le deuxième progrès important à attendre est la capacité à découvrir les liens de causalité. Un algorithme actuel est incapable d'organiser un système logique sur les données, de le tester et de le valider, il est incapable de construire seul des inférences causales. On désigne par inférence le fait de deviner quel facteur (par exemple, genre, formation, etc.) influe sur quel autre facteur (par exemple, salaire, promotion, etc.). L'IA actuelle aide à confirmer de telles relations, sans autonomie pour imaginer et prouver la causalité.

Une fois capable de détecter automatiquement les catégories symboliques et leurs liens de causalité, il sera alors possible de les manipuler avec les anciennes techniques développées pour les systèmes experts. Celles-ci permettent la récursivité, la capacité de faire des inférences sur des inférences, c'est-à-dire raisonner sur un raisonnement. Le troisième progrès fondamental sera donc la jonction des systèmes experts avec les techniques de ML, joignant la symbolique et la récursivité de ces anciennes techniques aux échelles calculatoires massives permises par les nouvelles techniques du ML.

## Un long hiver à venir, en attendant que la machine trouve sa motivation

En attendant de tels progrès, seuls les humains savent construire du savoir, c'est-à-dire choisir quoi analyser, faire des hypothèses, les vérifier, etc. Au mieux, ils peuvent utiliser l'IA actuelle comme un assistant statistique sophistiqué.

Pour que l'IA puisse contribuer avec une relative autonomie au processus d'élaboration du savoir, les progrès techniques requis sont énormes. Tout d'abord, ceci nécessitera des volumes de calculs phénoménaux au regard des capacités actuelles. Pour comparaison, les cerveaux biologiques sont de plusieurs ordres de grandeur plus efficaces que le silicium, à la fois en volumétrie de calcul et en consommation énergétique. Mais la loi de Moore n'a pour l'instant jamais failli, et il est donc à parier que les capacités de calculs vont continuer à monter de manière toujours plus étonnante.

De manière plus cruciale, ces calculs récursifs doivent être conduits en permanence, et chaque calcul ne peut être conduit « à l'infini » sur chaque élément. Il va donc falloir inventer une informatique fondée sur les arbitrages d'apprentissage : non seulement la

machine décidera toute seule d'initier une recherche d'inférences, mais elle doit aussi savoir s'arrêter et se contenter d'un bon modèle qui fonctionne. Et elle saura aussi de reprendre quand elle détecte qu'elle doit apprendre de nouveau.

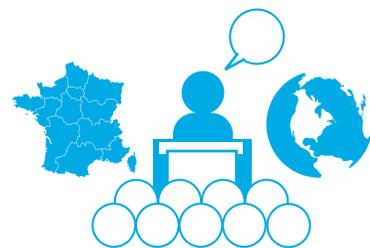
Probablement que ce processus ressemblera à ce que nous appelons, dans le système cognitif humain, de la « motivation ». En effet, l'humain apprend en permanence de son environnement, la motivation étant un mécanisme crucial dans le choix entre Agir et Apprendre. Par contraste, la machine est pour l'instant tellement simple que son maître décide pour elle quand — et quoi — calculer. Comme pour l'humanité, l'intelligence n'apparaîtra chez les machines que consubstantiellement à une forme de libre arbitre. Pour l'instant, la modélisation de ces formes d'émergence, de cognitions motivées, n'est pas vraiment commencée.

Globalement, rien n'interdit d'imaginer la mise au point d'IA beaucoup plus souple et conviviale que les algorithmes étroits d'aujourd'hui. Néanmoins, il est à parier qu'il faudra attendre longtemps, qu'il s'écoulera au moins un long hiver, ou même plusieurs hivers successifs, avant d'atteindre ce point mythique où la machine est d'une intelligence comparable à l'humain. Et étonnamment, cette intelligence artificielle, comme l'intelligence humaine, n'émergera qu'une fois adossée à des ressorts motivationnels solides. ■

*Publié initialement le 17 septembre 2020*

# LA RECHERCHE DEPUIS 2015

---



**816**  
ARTICLES PRÉSENTÉS  
À DES CONFÉRENCES SCIENTIFIQUES  
EN FRANCE ET À L'INTERNATIONALE



**ESSEC**  
BUSINESS SCHOOL

■ **ESSEC Business School**

3 avenue Bernard-Hirsch  
CS 50105 Cergy  
95021 Cergy-Pontoise Cedex  
France  
Tél. +33 (0)1 34 43 30 00  
[www.essec.edu](http://www.essec.edu)

■ **ESSEC Executive Education**

CNIT BP 230  
92053 Paris-La Défense  
France  
Tél. +33 (0)1 46 92 49 00  
[www.executive-education.essec.edu](http://www.executive-education.essec.edu)

■ **ESSEC Asia-Pacific**

5 Nepal Park  
Singapour 139408  
Tél. +65 6884 9780  
[www.essec.edu/asia](http://www.essec.edu/asia)

■ **ESSEC Africa**

Plage des Nations - Golf City  
Route de Kénitra  
Sidi Bouknadel (Rabat-Salé)  
Maroc  
Tél. +212 (0)5 37 82 40 00  
[www.essec.edu](http://www.essec.edu)

■ **Relations presse**

Vitaline GOMES  
Tél. +33 (0)1 34 43 30 29

■ **Rédactrice-en-chef**

Julia SMITH  
Tél. + 33 (0)1 34 43 98 25