INTELLIGENCE ARTIFICIELLE : OPPORTUNITÉS ET ENJEUX MANAGÉRIAUX[1]

Frédéric MARTY

CNRS – GREDEG – Université Côte d'Azur

Résumé

Si le recours à l'intelligence artificielle pour les algorithmes de prix, de recherche ou encore d'appariement génère des gains d'efficience dont bénéficient en premier lieu les consommateurs, les firmes doivent être conscientes que ces algorithmes peuvent générer des situations de non-conformité vis-à-vis des règles de protection de la concurrence et du consommateur et susciter dès lors qu'ils peuvent exposer à des risques réputationnels significatifs dès lors que leurs résultats sont vus comme de nature à restreindre, à manipuler les choix des consommateurs voire à introduire des discriminations. Cette contribution se propose de caractériser ces risques et insiste sur l'intérêt pour les entreprises à mettre en œuvre des politiques de conformité pour prévenir ces dommages ou y mettre terme rapidement et efficacement au travers d'audit algorithmiques.

Mots clés : algorithmes, intelligence artificielle, manipulation du consommateur, pratiques anticoncurrentielles, politiques de conformité, audits algorithmiques

Codes JEL : K21, K13

Abstract

While the use of artificial intelligence for pricing, search or matching algorithms generates efficiency gains that primarily benefit consumers, firms must be aware that these algorithms can generate situations of non-compliance with competition and consumer protection rules, and that they can expose them to significant reputational risks if their results are perceived as restricting or manipulating consumer choices or even as leading to discriminatory practices. This contribution aims to characterize these risks and insists on the need for companies to implement compliance policies to prevent these damages or to put an end to them quickly and efficiently through algorithmic audits.

Keywords: algorithms, artificial intelligence, consumer manipulation, anticompetitive practices, compliance programmes, algorithmic audits

JEL Codes: K21, K13

L'intelligence artificielle (ci-après IA) est indubitablement porteuse de gains d'efficience. Les consommateurs bénéficient d'ores et déjà de ces derniers. Pour autant, si l'IA est porteuse d'opportunités pour les firmes, elle peut conduire à des résultats qui peuvent porter préjudice à la protection de la concurrence et des consommateurs (B). Il est donc nécessaire d'envisager les mesures que peuvent prendre les entreprises pour faire face aux risques induits tant en matière légale que réputationnelle (C).

A – L'IA porteuse de gains d'efficience pour les firmes, leurs partenaires commerciaux et leurs clients

L'intelligence artificielle est avant toute chose un outil de prédiction algorithmique. Pour les firmes, ces capacités de prédiction sont des vecteurs essentiels d'optimisation des processus industriels et des chaînes logistiques. Dans le domaine de la logistique, l'IA est utilisée par les grandes places de marché pour prépositionner les stocks et pour optimiser l'organisation des entrepôts. Dans le domaine productif, l'IA est cruciale pour le développement d'une industrie 4.0 et constitue un vecteur essentiel, avec les blockchains, de croissance du modèle des plateformes industrielles. Elle permet notamment de raccourcir très significativement les boucles de rétroaction dans les activités de production et donc d'ajuster les caractéristiques des produits aux retours d'expérience tirés des premières unités produites et mises sur le marché. L'IA permet un suivi en temps du fonctionnement des produits et des systèmes et ouvre donc également à la mise en œuvre de stratégies de maintenance prédictives, particulièrement génératrices de gains d'efficacité. Enfin, toujours dans le domaine de la production, l'IA permet un meilleur ajustement de la production à la demande, dimension déterminante dans certaines industries où le stockage est coûteux sinon impossible. L'utilisation de l'IA dans la production et le pilotage de systèmes électriques est un exemple d'un secteur dans lequel les capacités prédictives sont génératrices de gains d'efficacité particulièrement significatifs[1].

Les gains liés à l'IA ne sont pas limités aux grandes entreprises du numérique, quand bien même elles y jouent un rôle déterminant. Les capacités de développement et les possibilités d'utiliser les ressources fournis par les services d'infonuagique (*cloud*) facilitent la mise en œuvre de solutions d'IA par les PME. Ces services permettent d'abord de réduire les barrières à l'entrée pour développer des solutions d'IA spécifiques à la firme. La plateforme peut offrir des capacités de stockage, de codage et de calcul. Elle permet également d'entraîner les algorithmes tout en fournissant des garanties quant à la confidentialité des données. Les firmes peuvent également utiliser des solutions d'IA, fournies par des entreprises de services ou par les plateformes elles-mêmes.

[1] Milgrom P.R. and Tadelis S., (2019), "How Artificial Intelligence and Machine Learning Can Impact Market Design", *in* Agrawal A., Gans J., Goldfard A., eds., *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, pp. 567-585.

Au-delà des gains d'efficience dans les processus industriels, l'IA est source d'un meilleur service pour les consommateurs.

Premièrement, l'IA joue un rôle prépondérant dans les algorithmes de recommandations et dans les algorithmes d'appariement (*matching*). Le recours à l'IA permet de mieux prédire les préférences et les besoins des consommateurs. Reliée à l'industrie 4.0, elle peut permettre d'ajuster les produits et les services proposés à chaque consommateur à ses besoins propres. Il est alors possible sur le principe de concilier production de masse et sur mesure. Il s'agit d'une logique *versioning* dans le bon sens du terme permettant une parfaite adéquation aux besoins du consommateur.

Dans le domaine des algorithmes de recherche, l'IA peut *a priori* affiner les résultats en fonction des caractéristiques du demandeur. Les algorithmes sans IA peuvent le faire dans un environnement où l'utilisateur est identifié et dans lequel un historique de navigation existe déjà. Au-delà de cette capacité de déduction, l'algorithme met en œuvre une micro-segmentation conduisant à une recommandation finement adaptée. Cela peut s'avérer favorable pour le consommateur (sous certaines conditions comme nous le verrons *infra*). L'algorithme exerce une fonction de filtrage. Non seulement, les coûts de recherche sont limités au profit du consommateur mais ce dernier serait soulagé d'être exonéré des effets de la tyrannie des choix. Cependant, il n'est pas acquis que le consommateur préfère réellement être directement aiguillé vers un résultat particulier dès la première étape de sa stratégie de recherche. L'optimisation des moteurs de recherche par l'IA peut au contraire accompagner le comportement de l'agent plus que le canaliser et le contraindre. C'est le sens de l'article de Blake, Nosko et Tadelis publié en 2016[2]. Un algorithme de recherche doit, s'il veut être performant, accompagner le comportement « naturel » de recherche de recherche des consommateurs. Il s'agit d'opérer une restriction progressive de l'éventail de choix et non pas d'imposer une solution par défaut. En effet, dans une démarche de recherche naturelle telle qu'observée, le consommateur commence par une décision d'exploration et termine par une décision d'exploitation (tout comme un algorithme d'IA d'ailleurs). La recherche est un processus de découverte et pas seulement une friction ou un coût. Elle joue un rôle dans la construction des préférences du consommateur –. Si elle est finalisée trop vite, si les frictions sont évitées, en conduisant le consommateur à opter pour un choix rapidement, cela peut réduire son utilité. En ce sens l'IA peut conduire à un

[2] Blake T., Nosko C., and Tadelis S., (2017), "Returns to Consumer Search: Evidence from eBay", *17th Conference on Electronic Commerce*, EC2016, pp.531-545

ajustement progressif des résultats de l'algorithme de recherche au comportement observé du consommateur.

Deuxièmement, l'IA peut jouer un rôle déterminant dans la fixation des prix. Outil de prédiction, l'IA peut permettre d'inférer la capacité à payer de chaque segment de consommateur… voire conduire à personnaliser les prix en établissant chacun au niveau de la capacité maximale à payer de chaque consommateur. L'individualisation des prix a quelques mérites en termes d'efficacité en ce qu'elle permet la mise en œuvre de subventions croisées entre les consommateurs. Elle conduit à un volume de transactions supérieur à l'équilibre que ce que permettrait un prix uniforme[3].

B – Un recours à l'IA porteur de risques pour l'entreprise

Si l'IA peut apporter maints avantages aux entreprises, elle peut également générer des risques légaux et réputationnels. Nous pouvons distinguer les risques d'infraction aux règles de concurrence et les risques reliés à la protection des consommateurs. Dans les deux cas, les enjeux pour les entreprises sont d'autant plus significatifs que les infractions peuvent être involontaires et difficiles à détecter y compris par les entreprises elles-mêmes. La première raison tient au fait que les algorithmes prennent automatiquement des décisions relatives aux prix qui ne tiennent pas à seul codage initial mais de leur apprentissage autonome. La seconde raison est liée au fonctionnement de type boîte noire de certains algorithmes. La prédiction qui est faite quant au prix à proposer à chaque consommateur n'est que très difficilement explicable. De surcroît, les prix en ligne se caractérisent par un foisonnement très significatif résultant du jeu même des différents algorithmes de prix. La détection de « décisions » algorithmiques problématiques, tant en regard des règles de concurrence qu'en regard de celles afférentes à la protection des consommateurs, notamment contre les discriminations.

Le recours à des algorithmes de prix utilisant un apprentissage autonome auto-renforçant peut générer des équilibres de collusion tacite sans que pour autant ces derniers soient codés pour parvenir à ce résultat[4]. Non seulement, les algorithmes peuvent converger spontanément vers un équilibre mutuellement avantageux mais ils peuvent le faire sans échange d'information ou encore transparence artificielle. Il n'est dès lors plus possible de fonder une sanction sur la présence de pratiques facilitatrices même si les autorités de concurrence détectent des

[3] Marty F., (2019), "Plateformes numériques, algorithmes et discrimination", *Revue de l'OFCE*, volume 164, 2019-4, pp.91-118

[4] De Marcellis-Warin N., Marty F. et Warin T., (2021), « Vers un virage algorithmique de la lutte anticartels ? Explicabilité et redevabilité à l'aube des algorithmes de surveillance », *Ethique Publique*, 23, 2-2021.

configurations de prix anormales au moyen d'outils algorithmiques de supervision des marchés[5]. De tels équilibres peuvent émerger dans des marchés dont les prix ne sont pas déterminés par algorithmes. Cependant, ils ne peuvent s'observer que dans des conditions de marché très simples, avec notamment peu d'acteurs et des produits très homogènes. Ce que peuvent changer les algorithmes de prix, c'est que ces équilibres pourraient émerger plus rapidement et dans des environnements de marché bien plus complexes.

Si ce scenario de collusion algorithmique spontanément générée par algorithmes a été critiqué dans la littérature comme recouvrant des cas extrêmement peu probables[6], de récents travaux tendent à montrer qu'ils n'en sont pas moins possibles. Premièrement, il a pu être montré par simulations numériques que des algorithmes de Q-learning mis en œuvre par des concurrents pouvaient rapidement conduire vers des équilibres collusifs sans échange d'informations ni transparence artificielle et que de surcroît ces équilibres étaient résilients, en d'autres termes qu'ils reviennent à une situation d'équilibre collusif après un choc exogène[7]. Au-delà de ce résultat expérimental, une seconde contribution empirique a pu relier la mise en œuvre d'algorithmes de Q-learning avec la convergence vers des configurations d'équilibre collusifs dans le secteur des stations d'autoroutes en Allemagne[8].

Ce scenario soulève des problèmes de responsabilité. La convergence vers l'équilibre collusif n'est pas liée à l'utilisation d'un même algorithme, ne correspond pas à un codage délibéré et n'est que la résultante que des décisions d'exploration et d'exploitation des algorithmes mis en œuvre par les concurrents. La collusion peut être vue comme la résultante de la compréhension de la concurrence par les firmes et de leurs absences de biais décisionnels les conduisant à se comporter spontanément dans le sens de la maximisation de la profitabilité de la firme. La firme peut-elle être tenue pour responsable en l'absence même de pratiques facilitatrices ? Le développeur peut-il être tenu comme responsable pour défaut de diligence dans l'appréciation des dommages concurrentiels que l'algorithme pourrait provoquer au travers de ses interactions avec des algorithmes tiers ? Est-ce un défaut de supervision ?

---

[5] Voir par exemple, Schrepel T. and Groza T., (2022), "The Adoption of Computational Antitrust by Agencies: 2021 Report", Stanford *Computational Antitrust*, 78(2) https://ssrn.com/abstract=4142225

[6] Ittoo A. et Petit N. (2017), "Algorithmic Pricing Agents and Tacit Collusion: A Technological Perspective" *in* Jacquemin H. et de Streel A. *L'intelligence artificielle et le droi*t, Larcier: 241-256.

[7] Calvano E., Calzolari G., Denicolò V., and Pastorello S., (2020), "Artificial Intelligence, Algorithmic Pricing, and Collusion", *American Economic Review*, 110 (10): 3267-97.

[8] Assad S., Clark R., Ershov D., and Xu L., (2020), "Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market", *CESifo Munich Working Paper*, n°8521,

A ces risques d'ententes entre firmes dans des conditions de marché qui ne sont pas seulement celles d'oligopoles étroits, il faut ajouter des risques de pratiques unilatérales mises en œuvre par des opérateurs dominants, eux aussi basés sur l'IA. Les firmes concernées sont alors généralement de grandes firmes. Cependant la dominance ne correspond pas en elle-même à des critères de taille. Elle s'apprécie par rapport à un marché pertinent dont les délimitations peuvent correspondre à un marché de niche. S'il n'existe guère de substitut au service fournis par une entreprise, aussi petite soit-elle, elle peut être considérée comme dominante[9].

Quelles peuvent être les pratiques en cause ? Les premières pratiques susceptibles d'être sanctionnées correspondent à des abus d'exploitation. Si l'algorithme permet de prédire le prix maximal que peut payer chaque client ou chaque partenaire commercial, l'entreprise est en position d'extraire tout le surplus de l'échange. En droit de la concurrence de l'UE, un abus d'exploitation peut correspondre non seulement à des prix « excessifs » mais également à des prix discriminants. Or, les algorithmes utilisant l'IA permettent d'individualiser les prix en fonction non pas du coût induit par la prestation mais de la prédiction de la propension maximale à payer de chaque consommateur. L'entreprise doit alors veiller à ce que ses pratiques ne soient pas sanctionnables sur cette base. Elle doit les auto-évaluer et ce faisant superviser le fonctionnement de ses algorithmes, lesquels peuvent spontanément conduire à ce type de pratiques.

Les secondes pratiques qui peuvent être sanctionnées sur la base d'un abus d'éviction. S'il est légitime (et efficace économiquement) qu'une entreprise en supplante une autre sur le marché, cela est conditionné par le fait que la sortie de l'entreprise considérée ne procède que des seuls mérites de l'entreprise dominante et non pas de l'instrumentalisation des avantages qui sont liés à sa position de marché. Les avantages algorithmiques que peut détenir une firme dominante sur un marché donné peuvent conduire à une sanction concurrentielle si ces derniers sont utilisés dans des conditions non conformes aux règles de concurrence. Par exemple, une extraction indue de données (qui peut être considérée comme abus d'exploitation) peut permettre d'entraîner une IA mieux que ne pourraient le faire les concurrents et donc être à la base d'un avantage algorithmique conduisant à une éviction des concurrents.

L'avantage algorithmique peut être à la source d'une éviction qui peut être qualifiée d'anticoncurrentielle. Cela peut correspondre aux pratiques dites d'auto-préférence (*self-*

[9] Salemme E., (2022), « La détermination du marché pertinent », *in* Mezaguer M., coord., *Le droit antitrust de l'Union européenne,* tome 1, Les dispositions générales, Commentaire J. Mégret, Institut d'Etudes Européennes, Editions de l'Université de Bruxelles, Bruxelles, juin, pp.277-295.

preferencing). Un moteur de recherche, un algorithme de recommandation ou d'appariement peut être biaisé en faveur d'un service donné[10]. Le service en question peut être un service fourni par l'entreprise dominante ou par un opérateur tiers. Dans le cas où il s'agit d'avantager son propre service, l'entreprise peut être accusée de vouloir étendre sa position dominante d'un segment de marché à un autre. On parle alors de stratégie de levier anticoncurrentiel. Quand il s'agit de privilégier une entreprise complémenteuse au détriment d'une autre, la motivation stratégique peut être de deux sortes. La première logique peut être de favoriser le complémenteur qui est le plus rémunérateur en termes de commissions ou de transfert de données. La seconde logique peut consister en l'entrave à l'accès au marché d'un complémenteur qui menacerait potentiellement la position dominante actuelle (au travers d'une possible intégration verticale) ou qui se montrerait insuffisamment coopératif (en refusant par exemple de n'être présent que sur la plateforme considérée). A nouveau, il appartient à la firme dominante de s'assurer que ses algorithmes ne conduisent pas à ce type de biais.

Il est également à relever qu'un avantage informationnel d'une firme dominante acquis indûment au détriment de ses partenaires commerciaux peut potentiellement lui permettre d'entraver la concurrence au travers de meilleures capacités de prédictions algorithmiques. C'est le cas notamment des stratégies de prédiction du présent (*now-casting*). La firme dominante peut identifier bien plus en amont que les tiers (concurrents, partenaires commerciaux, autorités de concurrence, …) les risques et opportunités stratégiques. Cet avantage peut permettre de restreindre préventivement l'accès au marché des opérateurs porteurs de menaces concurrentielles, de cloner éventuellement leurs produits et services voire de les acquérir préventivement (parfois avant même qu'ils n'accèdent au marché). Ces stratégies sont décrites dans la littérature économique comme des stratégies de zones mortifères (*kill zones*) quand il s'agit d'éviction anticoncurrentielle ou d'acquisitions tueuses (*killer acquisitions*) ou du moins consolidantes quand il s'agit de stratégies passant par des rachats d'entreprises[11].

La loi européenne sur les marchés numériques (*Digital Markets Act*) adoptée en mai 2022 repose en grande partie sur la prise en compte de ces enjeux et de la capacité des firmes désignées comme contrôleuses d'accès (*gatekeepers*) d'entraver la concurrence dans les écosystèmes numériques. Si le DMA ne concerne par définition que de très grandes entreprises

[10] Bougette P., Gautier A., and Marty F., (2022), "Business Models and Incentives: For an Effects-Based Approach of Self-Preferencing?", *Journal of Competition Law and Practice*, 13(2), March, pp.136-143.
[11] Harnay S., Marty F. et Toledano J., (2019), « Algorithmes et décision concurrentielle : risques et opportunités », *Revue d'Economie Industrielle*, n° 166, pp. 91-118.

de plateformes, les règles de concurrence continuent à s'appliquer à toute entreprise. La sanction des abus d'éviction et d'exploitation peut notamment s'appliquer comme nous l'avons supra à toute entreprise dominante sur un marché pertinent donné.

En matière de risques juridiques, les entreprises utilisatrices d'IA ne sont pas seulement exposées aux règles afférentes à la protection de la concurrence ou de la loyauté des relations commerciales mais également à celles relatives à la protection du consommateur. La personnalisation des recommandations algorithmiques et leur ajustement immédiat aux décisions et aux comportements observés des consommateurs peuvent conduire à une limitation indue de l'espace de leurs choix voire à une manipulation de ces derniers[12].

Il peut s'agir de deux pratiques distinctes ; les premières correspondent au cadre habituel des pratiques commerciales trompeuses ou mensongères, les secondes à des architectures de choix sinon trompeuses du moins biaisées. Il s'agit des *dark patterns*, c'est-à-dire des modes de présentations de choix qui poussent les utilisateurs à agir contre leurs intérêts (*bad nudges*) ou qui les empêchent d'agir dans ce sens, ou du moins les entravent (*bad sludges*). Ces stratégies ne sont pas spécifiques à l'IA ni au numérique[13]. Hors de la sphère de la concurrence, elles ont été observées en matière de protection des données personnelles et de recueil du consentement dans le cadre du RGPD[14]. Dans le domaine des activités de marché, les capacités de mise en œuvre de ces pratiques et leur efficacité potentielles sont démultipliées. L'IA peut *augmenter* ces pratiques en les personnalisant et en les rendant dynamiques. Elle peut également les mettre à la portée de nombreuses entreprises et pas simplement des opérateurs dominants.

Les algorithmes à l'œuvre sont alors indubitablement profitables pour la firme et possiblement efficacement économiquement en termes de bien-être agrégé. Cependant, la performance des algorithmes peut induire des dommages aux consommateurs. Cet impact est bien plus direct et

---

[12] De Marcellis-Warin N., Marty F., Thelisson E., and Warin T., (2022), "Artificial intelligence and consumer manipulations: from consumer's counter algorithms to firm's self-regulation tools, *AI & Ethics*, 2(2), May, pp.259-268.

[13] Le biais peut, par exemple, être volontairement produit par le design même d'une page Internet qui rend par exemple difficile la comparaison des prix des produits concurrents en un coup d'œil.
Johnen J. and Tsz Kin Leung B., (2022), "Distracted from Comparisons: Product Design and Advertisement with Limited Attention", *LIDAM Discussion Paper CORE*, 2022/17, April, 65p.
Le biais peut également procéder d'une révélation volontairement tardive du prix d'un produit au travers de la révélation progressive des frais annexes. Il s'agit également dans ce cas d'entraver la capacité qu'a le consommateur à comparer les prix.
Johnen J. and Somogy R., (2022), "Deceptive Features on Platforms", *LIDAM Discussion Paper CORE*, 2022/19, April, 64p.

[14] Gray C.M., Santos C., Bielova N., Toth M., and Clifford D., (2021), "Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective", *CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, May, art. 121, pp.1-18, https://doi.org/10.1145/3411764.3445779

significatif que le cas de la publicité traditionnelle. La « construction des préférences » se fait à une autre échelle. Ainsi, la prédiction algorithmique victime de son succès. La personnalisation des prix qui permet de placer son prix au niveau de la propension maximale à payer du consommateur peut être analysée comme une extorsion de son surplus. La personnalisation des offres (c'est-à-dire la face sombre du *versioning*) peut conduire non pas à un ajustement de la qualité du produit aux besoins du client mais un ajustement à son niveau d'expertise. Un même prix peut être appliqué à deux clients, mais le niveau de qualité sera moindre pour le « naïf » et la marge supérieure pour la firme. Il ne faut jamais oublier qu'un prix et qu'une qualité uniforme jouent au moins partiellement comme une protection des consommateurs les moins informées par les consommateurs mieux informés.

La discrimination qui résulte du fonctionnement des algorithmes peut entraîner des risques légaux et économiques pour les firmes. Premièrement, il peut s'agir d'une discrimination pour les consommateurs en matière d'accès au marché des biens et services et aux plus généralement aux opportunités offertes par le marché du travail ou encore le marché du crédit. Ces discriminations peuvent engager leur responsabilité. Deuxièmement, l'entreprise fait face à un risque réputationnel tant vis-à-vis de ses consommateurs (*consumer backlash*) que des autres parties prenantes, qu'il s'agisse des partenaires commerciaux, salariés ou financeurs. Un enjeu d'éthique est à prendre en considération et il ne se limite pas à une question de responsabilité (ou de management d'un risque juridique) mais est déterminant dans l'engagement des parties prenantes dans la vie et dans la dynamique de l'entreprise, c'est-à-dire dans sa capacité à générer de la valeur.

C - Quelle étendue du problème pour les firmes ?

Comme nous l'avons relevé *supra* deux dimensions sont à prendre en compte pour les entreprises. La première est légale. Il s'agit alors d'appréhender les effets de l'IA en termes de risques et donc de responsabilité de la firme. La seconde est la réputationnelle. Le coût peut alors être important que la firme a pris des engagements éthiques vis-à-vis de valeurs sociétales. Une discrimination observée ex post et résultant du jeu spontané des algorithmes peut avoir un effet équivalent à des accusations d'écoblanchiment. Le problème lié à l'IA est que les pratiques en cause peuvent être en partie involontaires. L'IA fonctionne comme une boîte noire. Les biais peuvent venir des données, du codage de l'apprentissage autonome.

Des expériences menées sur des agents conversationnels montrent que par leur apprentissage ceux-ci peuvent développer des biais inattendus et particulièrement préjudiciables pour l'image de la firme. De la même façon des algorithmes utilisés pour l'accès aux crédits peuvent s'avérer biaisés par l'utilisation de *proxies* tels les codes postaux dans l'appréciation des risques de non-remboursement qui peuvent soit conduire à des refus de prêts soit conduire à des taux plus élevés et donc conduire à un plus fort taux de défaillance des emprunteurs. L'effet sera d'autant plus préjudiciable qu'une boucle de rétroaction se constituera. Les prédictions se confirmant, le biais sera renforcé dans les décisions futures. En d'autres termes, la prédiction sera autoréalisatrice. L'absence de supervision de ce qui devrait rester des recommandations algorithmiques – et non des décisions algorithmiques – peut donc confirmer et aggraver des biais sociétaux.

Le risque est ici d'avoir un phénomène de *man out the loop*, c'est-à-dire d'absence de supervision ou de contrôle du fonctionnement des algorithmes. La déresponsabilisation des agents peut se solder par l'engagement de la responsabilité de la firme. Une seconde variante de cette déresponsabilisation peut provenir d'un comportement rationnel des agents. Ils peuvent suivre « mécaniquement » les recommandations de façon à ne pas engager leur responsabilité. S'écarter de la recommandation algorithmique expose à devoir se justifier en cas de difficultés *ex post*. Par exemple, dans le cadre de l'octroi d'un crédit, s'aligner à une recommandation de refus ne soumet pas l'agent au risque d'expliquer son choix en cas de défaut de remboursement. L'alignement sur la recommandation rend de surcroît le scenario alternatif inobservable. Se placer *out of the loop* peut être rationnel pour l'agent… mais préjudiciable pour l'entreprise.

Il est également à noter que l'entreprise est d'autant plus exposée que les biais peuvent provenir des algorithmes utilisés qui sont développés et entraînés par des tiers et plus généralement des données sur lesquelles ils ont été entraînés. Ce faisant, le management des risques ne se limite pas à la formation et à la supervision de ses propres agents. Il est nécessaire qu'elle évalue les risques liés à la conception, au fonctionnement et à l'évolution de ses algorithmes.

Le premier point à prendre en considération tient à la transparence de l'algorithme ne suffit pas et est de fait illusoire du fait du fonctionnement de type boîte noire de l'apprentissage machine. Il faut qu'il soit sinon interprétable ou redevable (*accountable*). Il est nécessaire dans une logique de conformité de développer des méthodes de rétro-ingénierie pour déterminer à partir de l'observation des entrées et des sorties quelles sont les principales variables explicatives des prédictions algorithmiques. Ces audits ne se limitent pas à une étape préalable à la mise sur le marché mais doivent être mis en œuvre durant toute la durée de son utilisation dans la mesure

où les prédictions (voire la structure du code dans certains modèles d'apprentissage) vont évoluer à partir des nouvelles données observées et des interactions avec d'autres algorithmes.

Ainsi même si l'algorithme est développé par un tiers, son entraînement puis son fonctionnement doit être audité par l'entreprise qui va l'utiliser. Elle doit s'assurer à la fois d'une conformité par conception et d'une évolution de l'algorithme compatible avec le cadre légal et ses valeurs éthiques. Globalement, l'entreprise doit veiller à l'acceptabilité sociale de l'algorithme tant *ex ante* que dans le cadre de sa mise en œuvre. La logique à l'œuvre vis-à-vis des parties prenantes est alors proche de celle connue dans le domaine des projets d'investissements sous la dénomination de *social licence to operate*[15]. S'assurer de l'acceptabilité sociale des algorithmes ne se limite pas à la phase de développement mais doit s'inscrire dans une démarche continue.

Au-delà des enjeux réputationnels, cette logique d'auto-évaluation est également essentielle dans le management du risque du juridique. Le caractère de boîte noire de l'algorithme ou l'absence d'intentionnalité ne sont pas des lignes de protection. Toute entreprise qui met en œuvre des algorithmes pouvant avoir un impact significatif sur les concurrents, les partenaires commerciaux, les consommateurs ou toute autre partie prenante doit s'engager dans une démarche de conformité.

Il ne s'agit pas seulement d'une une auto-régulation basée sur une démarche éthique et responsable de l'entreprise mais également du management d'un risque légal et régulatoire. Les entreprises sont placées dans un cadre de régulation déléguée. Les coûts de la régulation sont à leur charge dans la mesure où elles peuvent éviter l'occurrence du dommage en dernier ressort et au moindre coût dans la mesure où elles peuvent jouer sur les incitations pour les développeurs et les utilisateurs à mettre en œuvre ex ante les précautions nécessaires pour prévenir les dommages (*due diligence*) et *ex post* à mettre fin rapidement à des fonctionnements des algorithmes qui pourraient s'avérer préjudiciables (*monitoring*). L'absence de mesures préventives contre des pratiques qui pourraient porter préjudice aux consommateurs peut conduire à mettre en cause la responsabilité des entreprises.

Un exemple distinct de notre propos mais assez représentatif peut être apporté par la procédure ouverte par la Federal Trade Commission américaine le 28 juin 2022 contre Walmart sur la base d'absence de mesures de contrôle et de prévention suffisantes (notamment en termes de

[15] Parsons R., Lacey J., and Moffat K., (2014), "Maintaining legitimacy of a contested practice: How the minerals industry understands its 'social licence to operate", *Resource Policy*, 41, September, pp.83-90.

sensibilisation des personnels) contre les transferts d'argent frauduleux, des virements par détournement des identifiants bancaires et autres usurpations[16]. Même si elle n'est pas responsable d'une pratique dommageable l'absence de diligence d'une entreprise pour détecter de telles pratiques et pour engager des mesures de protection effectives du consommateur peuvent lui être reprochées. Il en va de même pour les dommages liés au fonctionnement des algorithmes.

Avant de conclure cette contribution, il est possible d'illustrer notre propos en nous appuyant sur un récent article de Renée Richardson-Gosline publiée en juin 2022 dans l'*Harvard Business Review*[17]. Cet article illustre la nécessaire responsabilisation des entreprises utilisant des algorithmes d'IA au travers d'une question que nous avons abordé dans notre première partie, celle de la réduction des frictions dans les parcours clients. Ces frictions sont souvent analysées comme des coûts de transactions. Les firmes veulent éviter les frictions dans le parcours en ligne des consommateurs et peuvent *a priori* améliorer le bien-être collectif par ce biais. Or, toute les *sludges* ne sont pas préjudiciables aux consommateurs. Elles peuvent dans une mesure les protéger contre des achats impulsifs, limiter les risques liés au surendettement ou encore faire obstacle à d'éventuelles assuétudes. Supprimer les caisses et le paiement direct dans un magasin (en recourant à la reconnaissance faciale) peut certes améliorer l'expérience client, limiter les coûts de transaction… mais aggraver les risques décrits *supra*. De la même façon, comme nous l'avons vu guider le consommateur vers le choix qui est prédit comme le meilleur pour lui peut le priver de la capacité à construire ses propres préférences au travers des connaissances qu'il peut développer au cours de son comportement de recherche.

Il ne s'agit pas seulement de protéger le consommateur contre les architectures de choix trompeuses, les manipulations en ligne… et contre lui-même mais également de préserver voire de renforcer son agentivité. Ainsi, le codage et la supervision du fonctionnement des algorithmes doivent prévenir les biais (et surtout éviter qu'ils soient amplifiés et moins aisément détectables du fait du fonctionnement même de l'IA) mais devraient développer des frictions dans une logique éthique. Il ne s'agit pas seulement de mettre en œuvre des audits algorithmiques pour évaluer les effets potentiels des *bad nudges* des *bad sluges* mais d'expérimenter – éventuelle dans une logique de *bac à sable* (*sandbox approach*) des

[16] Federal Trade Commission, "FTC Sues Walmart for Facilitating Money Transfer Fraud That Fleeced Customers Out of Hundred of Millions", Press Release, 28 June 2022.
[17] Richardson Gosline R., (2022), "Why AI Customer Journeys Need More Frictions", *Harvard Business Review*, 9 June

architectures de choix éthiques de nature à augmenter le consommateur dans l'exercice de sa liberté de choix.

Au final, l'audit et la supervision des algorithmes sont stratégiques pour les entreprises elles-mêmes. Ils sont une réponse aux futures régulations de l'IA en cours de préparation tant au sein de l'UE qu'aux Etats-Unis. Ils s'inscrivent dans les engagements éthiques des entreprises et peuvent s'intégrer dans des codes de conduite unilatéraux et volontaires. Ils répondent enfin à un intérêt bien compris. Si la concurrence se fait par la qualité alors la protection du consommateur, notamment par la protection des données personnelles, de la confidentialité de son comportement en ligne et l'engagement contre les biais et les manipulations seront des paramètres déterminants pour le succès des firmes.