Points focaux pour une IA vérifiable et explicable

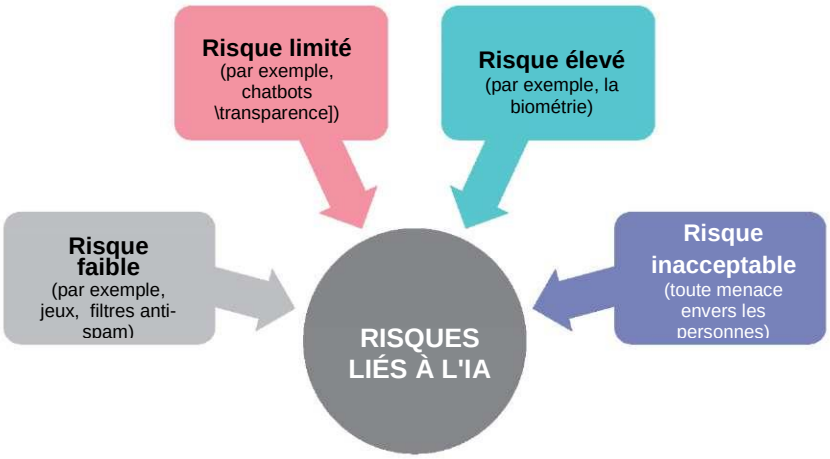
L'intelligence artificielle (IA) est une branche de l'informatique qui a vu le jour, du moins au niveau académique, il y a près de 75 ans. En général, elle s'intéresse aux machines informatiques intelligentes qui exécutent les types de tâches qui nécessitent l'intelligence humaine.

Le projet de Loi sur l'intelligence artificielle de l'Union européenne (EU AIA) vise à garantir que l'IA fonctionne et soit bénéfique à la société.[1] Il commence également à qualifier le risque que les systèmes d'IA présentent pour les personnes et la société et suggère quels types d'IA nécessiteront le plus de surveillance.

Le projet de règlement établit une distinction entre les IA présentant un risque inacceptable, un risque élevé et un risque faible ou limité **(figure 1).**[2] Le risque inacceptable en matière d' IA consiste en des déploiements qui constituent une menace pour la sécurité, les moyens de subsistance et les droits des personnes.[3] Les déploiements de l'IA dans les domaines de la biométrie, des infrastructures critiques, de l'éducation, de l'emploi, des services, de l'application de la loi, des migrations humaines et de la justice sont considérés comme à haut risque.[4]

Il est nécessaire de disposer d'une IA vérifiable et explicable , en particulier dans les applications à risque élevé et limité , afin de garantir qu'aucun préjudice n'est causé aux personnes et à la société.

FIGURE 1
Types de risques liés à l'IA



Présentation de l'IA statistique

Il existe deux branches de l'IA : symbolique et statistique (c'est-à-dire connexionniste). L'IA statistique, la plus récente des deux, est à la fois plus difficile à contrôler et à expliquer. L'IA symbolique est basée sur des règles, et les règles sont beaucoup plus faciles à expliquer et à contrôler que les algorithmes basés sur des données, qui peuvent être sujets à d'importantes valeurs aberrantes, tant du côté de l'entrée des données que du côté de la sortie de l'IA.

L'IA statistique est une approche ascendante de l'IA, dont bon nombre des méthodes ont été développées par des statisticiens.[5] Elle dépend de grands volumes de données pour entraîner les modèles d'IA. Les outils généraux d'IA statistique comprennent les algorithmes classiques d'apprentissage automatique et les réseaux neuronaux et les outils spécifiques comprennent la vision par ordinateur et le traitement du langage naturel (NLP).

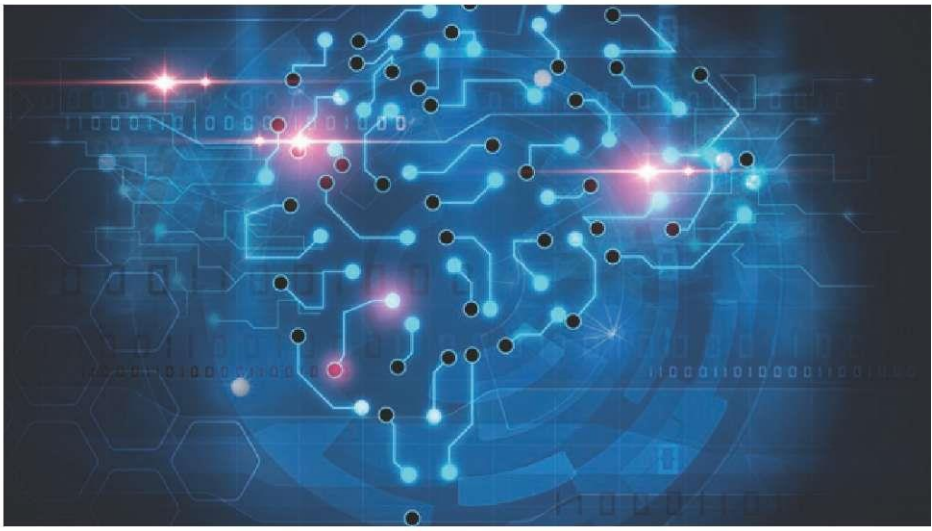
Parmi les exemples d'IA statistique, citons les recommandations de vente et les voitures à conduite autonome. Les techniques appliquées aux données comprennent la régression et la classification. L'apprentissage machine (un sous-ensemble de l'IA et de l'IA statistique) utilise principalement des techniques de réseaux neuronaux.

La présente discussion porte spécifiquement et implicitement sur la vérifiabilité et l'explicabilité de

GUY PEARCE | CGEIT, CDPSE

A une formation universitaire en informatique, en science des données et en affaires, et a occupé des fonctions de leadership stratégique, de gouvernance informatique et de gouvernance d'entreprise dans divers secteurs. Il est actif dans le domaine de la transformation numérique depuis 1999, en se concentrant sur l'intégration des personnes et des processus des technologies émergentes dans l'organisation afin d'assurer leur adoption effective. Son premier contact avec l'IA remonte à 1989 et il a suivi l'évolution de la discipline, de l'IA symbolique à l'IA statistique (connexionniste) au cours des décennies qui ont suivi. Il s'est vu décerner le prix 2019 ISACA® Michael Cangemi du meilleur auteur pour ses contributions à la gouvernance informatique et ses conseils en matière de transformation numérique, de risque, de gouvernance des données et de gouvernance informatique.

l'IA statistique, étant donné les défis que représente la réalisation de la validation dans cette branche de l'IA.



Carte conceptuelle des fonctionnalités de l'IA

L'IA statistique et l'IA symbolique se situent sur le même continuum d'IA (représenté par l'axe des y dans la **figure 2).** Vers l'extrémité statistique du continuum se trouvent les recommandations de vente de l'IA et vers l'extrémité symbolique les simulateurs d'échecs.

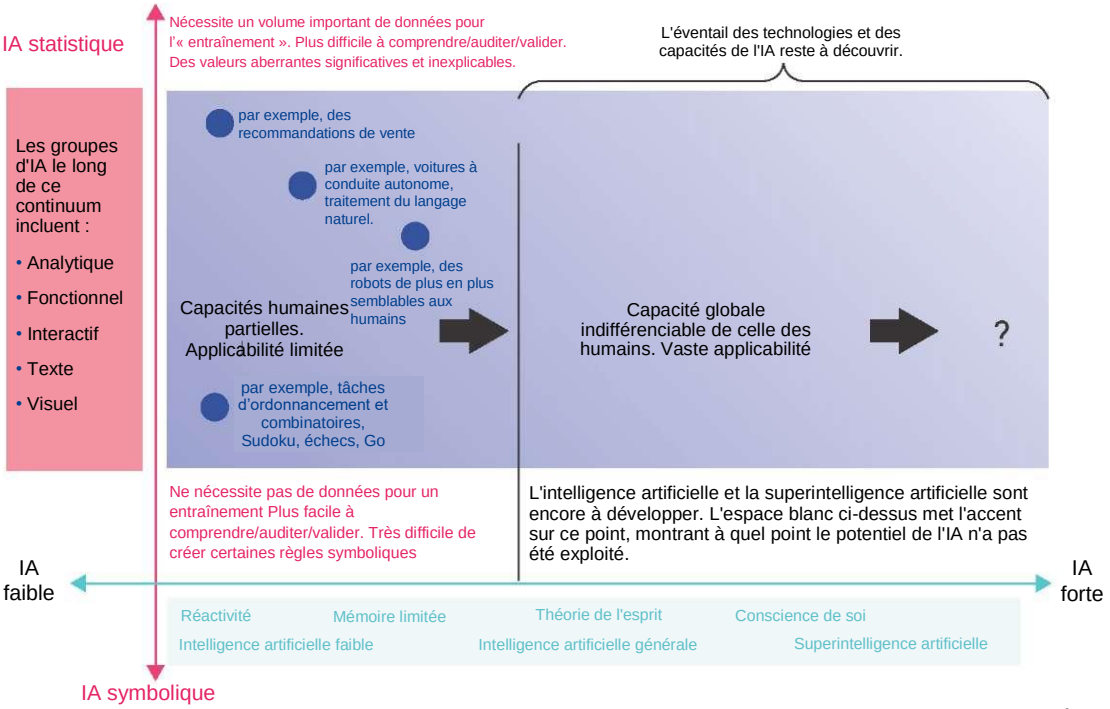
Il existe au moins un autre continuum classifiant l'IA, à savoir le continuum IA faible / IA forte (représenté par l'axe des x dans la **figure 2).** Le continuum faible /fort permet d'indiquer l'ampleur du travail de développement encore nécessaire pour amener l'IA au point où elle est intelligente (c'est-à-dire généralement capable d'exécuter des tâches d'une manière qui ne peut être distinguée de la manière dont un humain les exécuterait).

Capacité d'un déploiement d'IA à survivre à un examen minutieux

Indépendamment de la question de savoir si un système d'IA statistique est faible ou fort, la question qui devrait intéresser vivement tout professionnel de la gouvernance informatique est si la technologie d'IA d'un fournisseur et son déploiement associé peuvent survivre à un examen minutieux. Par exemple, quels sont les mécanismes disponibles pour évaluer la qualité et la pertinence des données d'apprentissage et déterminer si les algorithmes fonctionnent comme prévu ? Qui veille à ce que les résultats de l'IA servent l'humanité ? De ce point de vue, il existe au moins sept domaines de l'IA qui méritent d'être examinés de près **(figure 3).**[6]

FIGURE 2

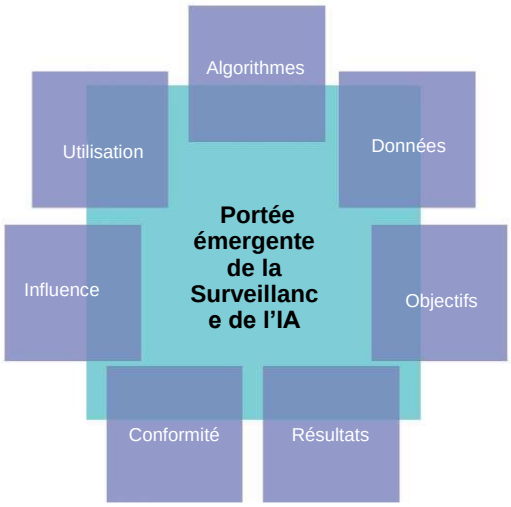
Le continuum IA faible-forte et le continuum IA statistique-symbolique



Source : Pearce, G.; "Real-World Data Resilience Demands an Integrated Approach to Artificial Intelligence, Data Governance and the Cloud," *ISACA® Journal,* vol. 3, 2022, https://www.isaca.org/archives

FIGURE 3

Domaines de l'IA qui nécessitent une surveillance



Source : Pearce, G.; M. Kotopski; « Algorithms and the Enterprise Governance of AI », *Journal de l'ISACA*, vol. 4, 2021, *https://www.isaca.org/archives*

Pour mettre en contexte les domaines de surveillance illustrés dans la **figure 3**, il est utile de comprendre que les données forment les algorithmes qui génèrent des résultats d'IA conformes aux objectifs de l'organisation pour la technologie. Les résultats : L'utilisation détermine la nature de l'influence des résultats comme étant éthiques ou non éthiques et affecte la présence ou l'absence de conformité légale. Une IA responsable exige que l'influence soit exercée pour le bien du public. L'utilisation fournit des informations pour la prise de décision par une personne ou un système automatisé. L'action entreprise sur cette information est ce qui génère finalement l'influence.

Les mêmes relations s'appliquent à l'IA symbolique, sauf concernant les données, car celles-ci ne sont pas nécessaires pour former l'IA symbolique. Tous les autres domaines requièrent la même surveillance que l'IA statistique, car chacune peut contribuer à l'IA en produisant des résultats inattendus. **La figure 4** présente des exemples de questions sur la surveillance pour chaque domaine de la **figure 3** afin de montrer la nature des contrôles et des équilibres nécessaires à l'IA.

L'éventail des défis de surveillance applicables à chaque domaine est important.

Par exemple, s'il y a des lacunes dans la surveillance des données, qui sont le fer de lance de la chaîne de valeur de l'IA statistique **(figure 5),** cela peut avoir un impact négatif sur l'ensemble du déploiement de l'IA et sur la qualité des actions prises sur la base des résultats de l'IA. Cela peut avoir un impact négatif sur les personnes, les communautés, les sociétés, les pays et même des régions géographiques entières pour lesquels l'IA est conçue pour produire des résultats.

Voici quelques exemples de caractéristiques importantes des données d'entrée à prendre en compte :

- **Qualité des données -** Si les données d'entrée sont de qualité inconnue ou médiocre, les résultats des systèmes d'IA seront certainement aussi de qualité inconnue ou médiocre. (Certaines dimensions de la qualité des données comprennent l'exactitude [c'est-à-dire le fait que les données se situent dans une fourchette attendue], la validité [c'est-à-dire, le fait que les données soient dans le format requis] et l'actualité [c'est-à-dire le fait que les données soient actuelles]). Les décisions basées sur l'IA peuvent tout simplement être totalement erronées. L'impact potentiel des organisations omettant les efforts de qualité des données pour l'IA est sérieux, avec la possibilité d'un impact négatif incommensurable sur les individus et même sur la société en général, sans parler de l'impact sur la réputation de l'organisation elle-même.
- **Volume de données -** Trop peu de données compromettent la qualité des résultats de l'IA, mais trop de données ne contribuent guère à améliorer la qualité du système d'IA et entraîneront des coûts supplémentaires et une complexité qui ne sont pas justifiables.

FIGURE 4

Exemples de questions concernant le Domaine de la surveillance de l'IA

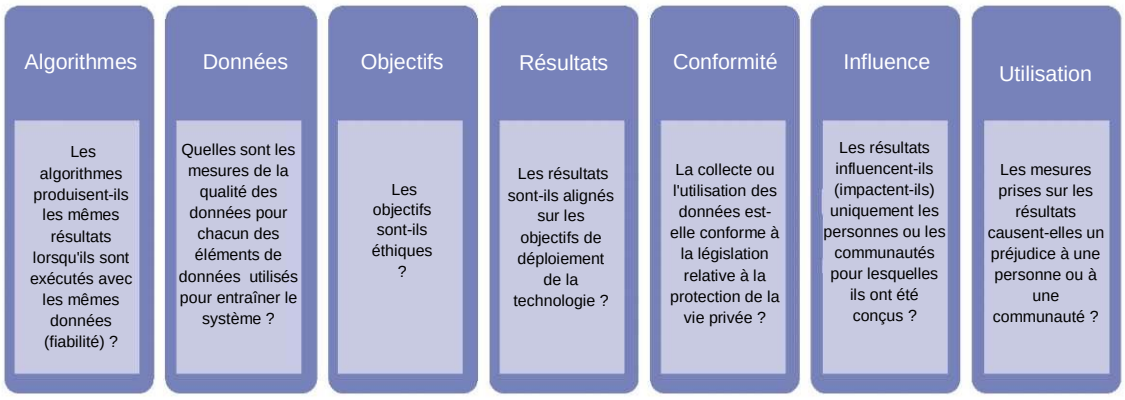
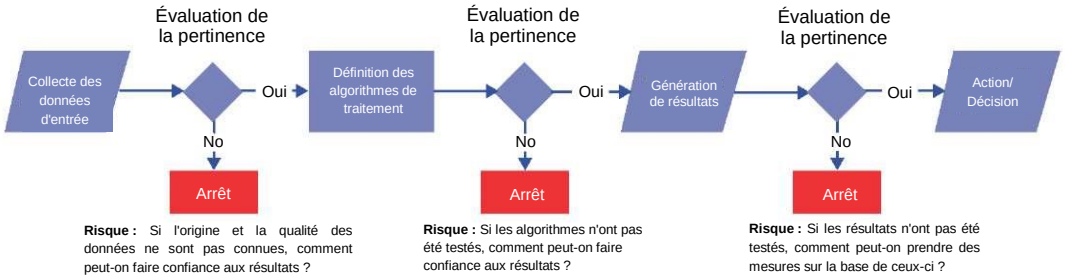


FIGURE 5

Chaîne de valeur simplifiée de l'IA



Source : Pearce, G.; « Data Auditing: Building Trust in Artificial Intelligence », *Journal de l'ISACA,* vol. 6, 2020, *https://www.isaca.org/archives*

- **Contenu des données (type de données) -** La nature des données d'entrée doit être compatible avec les objectifs du déploiement de l'IA . La sélection appropriée des données utilisées pour entraîner les algorithmes est essentielle au succès.
- **Dérive des données -** Des influences externes peuvent avoir un impact sur les données utilisées dans les modèles statistiques d'IA. Par exemple, le passage au travail à domicile pendant la pandémie de COVID-19 a influencé les données relatives au trafic, l'utilisation des transports publics et même la fréquentation des commerces. Si l'une de ces données a été utilisée pour former un modèle d'IA avant la pandémie, le modèle ne pourra pas être utilisé pendant la pandémie et probablement aussi après. Le modèle devra être réentraîné, avec des données provenant d'un délai spécifié dans ce cas, pour garantir la fiabilité des résultats du modèle.

« L'impact potentiel des organisations omettant les efforts de qualité des données pour l'IA est sérieux, avec la possibilité d'un impact négatif incommensurable sur les individus et même la société en général. »

Quant aux éléments de surveillance applicables aux étapes de la chaîne de valeur de l'IA, la **figure 6** illustre l'alignement des étapes entre la **figure 3** et la **figure 5.**

Performance en matière de surveillance organisationnelle

Quelle est la performance des organisations en matière de surveillance de l'IA ? Comme l'illustre la **figure 7**, pas très bonne.[7]

Il apparaît que de nombreux conseils d'administration (CA) ne sont pas prêts pour la transformation numérique.[8] Étant donné que l'IA est une technologie de transformation numérique, il s'ensuit que dans de nombreuses organisations, ni les CA ni la direction ne sont équipés pour assurer la surveillance de l'IA. De nombreuses organisations semblent conscientes que leurs déploiements d'IA

FIGURE 6

Relation entre la chaîne de valeur de l'IA et les domaines de surveillance de l'IA

Élément de la chaîne de valeur de l'IA statistique	Éléments pertinents de la surveillance de l'IA (d'après la Figure 3)
Collecte des données d'entrée	Données, objectifs
Définition des algorithmes de traitement	Algorithmes, objectifs
Génération de résultats	Résultats, conformité, utilisation
Action/prise de décisions	Influence, usage

FIGURE 7

Relation entre la gouvernance et la gestion de l'IA et les Domaines de surveillance de l'IA

Statistiques sur la gouvernance et la gestion de l'IA[a]	Éléments pertinents de la surveillance de l'IA (d'après la Figure 3)
Soixante-cinq pour cent des organisations ne peuvent pas expliquer comment fonctionne leur IA.	Algorithmes, données, résultats, utilisation
Soixante-treize pour cent des organisations ont du mal à donner la priorité à l'éthique et à l'IA responsable.	Conformité, influence, utilisation
Quatre-vingt pour cent des organisations ne contrôlent pas activement l'équité et l'éthique de leurs modèles de production.	Résultats, conformité, influence, utilisation
Soixante et un pour cent des membres des conseils d'administration ont une compréhension incomplète de l'éthique de l'IA.	Résultats, influence, utilisation
Soixante-sept pour cent des cadres supérieurs ont une compréhension incomplète de l'éthique de l'IA.	Influence, usage

Source : (a) Zoldi, S. ; « It's 2021. Do You Know What Your AI Is Doing? » FICO Blog, 25 mai 2021, *https://www.fico.com/blogs/its-2021-do-you-know-what-your-ai-doing*

ne survivront pas à un examen minutieux. Ces résultats devraient être troublants pour les professionnels de la gouvernance informatique. La gouvernance de l'IA, y compris la supervision efficace des algorithmes d'IA, est le meilleur instrument disponible[9] pour protéger l'organisation, ses clients et même la société dans son ensemble contre une IA irresponsable.[10]
Les bonnes intentions ne suffisent pas.

La nécessité de la vérifiabilité et de l'explicabilité de l'IA est claire, et toutes deux sont soumises à des attentes réglementaires croissantes. Par exemple, le Règlement général sur la protection des données (RGPD) de l'UE exige la vérifiabilité et l'explicabilité dans le traitement des données personnelles, y compris le traitement statistique de l'IA. La conformité est exigée de toutes les organisations qui offrent des biens ou des services à des clients ou à des entreprises de l'UE, qu'elles opèrent au sein ou en dehors de l'UE.

Du point de vue de la gouvernance d'entreprise, l'IA est tenue de respecter toutes les exigences légales applicables à l'organisation. Les organismes de services financiers, en particulier, « devraient revoir leurs politiques internes, leurs cadres de gouvernance et leurs pratiques contractuelles pour s'assurer qu'ils s'alignent sur les dernières réflexions concernant l'utilisation de l'IA. »[11]

Vérifiabilité de l'IA

L'IA vérifiable est une IA qui produit la documentation requise pour appuyer un examen réglementaire. Elle peut contribuer à atténuer les coûts juridiques potentiels, les atteintes à la réputation et le mécontentement des clients souvent associés aux processus non standard et aux décisions et résultats non documentés caractéristiques de nombreux modèles d'IA actuellement en production.[12] Il existe de multiples domaines qui nécessitent la vérifiabilité de l'IA. [13], [14]

FIGURE 8

Relations entre les domaines d'audit de l'IA et la surveillance de l'IA

Domaines d'audit [a,b]	Éléments pertinents de la surveillance de l'IA (d'après la Figure 3)
Gouvernance et gestion des projets d'IA	Objectifs
Données utilisées pour l'IA	Données
Approvisionnement en IA, conception et développement de modèles	Algorithmes
Outils d'IA en production	Résultats, utilisation
Évaluation de l'IA	Résultats, conformité, influence, utilisation

Sources : (a) Scanlon, L.; « Auditability of AI Vital for Financial Services », Out-Law Analysis, Pinsent Masons, 15 février 2021, *https://www.pinsentmasons.com/out-law/analysis/auditability-of-ai-financial-services;*
(b) Institutions supérieures de contrôle de la Finlande, de l'Allemagne, des Pays-Bas, de la Norvège et du Royaume-Uni, « Auditability Checklist », 24 novembre 2020, https://www.auditingalgorithms.net/AuditabilityChecklist.html

En d'autres termes, la vérifiabilité n'intervient pas seulement après la mise en production d'un système d'IA. L'ensemble du processus d'IA - de la planification aux exigences en matière de données, en passant par l'approvisionnement, le développement, la production et l'évaluation - doit être contrôlable. Non seulement le champ d'application de l'audit de l'IA est plus vaste que certains ne le pensent, mais de nombreux rôles doivent être consultés, comme le propriétaire du projet, le propriétaire du produit, l'utilisateur, le fournisseur d'assistance à l'utilisateur, le directeur de l'information (CIO), l'ingénieur des données, le développeur, le directeur de la sécurité de l'information (CISO), le directeur de la confidentialité (CPO) et le directeur financier (CFO).[15]
Il est important de noter que l'audit requiert la répétabilité du processus opérationnel, ce qui signifie que l'algorithme et les données brutes qui produisent un résultat spécifique de l'IA doivent être accessibles et disponibles. L'exigence de données brutes signifie que des processus de soutien doivent être mis en place pour garantir que les données brutes applicables à un résultat spécifique sont disponibles et peuvent être exécutées dans l'algorithme pour tester le résultat pendant l'audit.
La vérifiabilité de l'IA « a le potentiel de catapulter l'adoption en permettant une IA transparente et digne de confiance ».[16] Associée aux avancées en matière d'explicabilité des modèles d'IA, la vérifiabilité offre une fenêtre sur la santé de l'IA d'une organisation. Mais qu'est-ce que l'IA transparente ? Les exigences d'une IA statistique transparente sont les suivantes :[17]

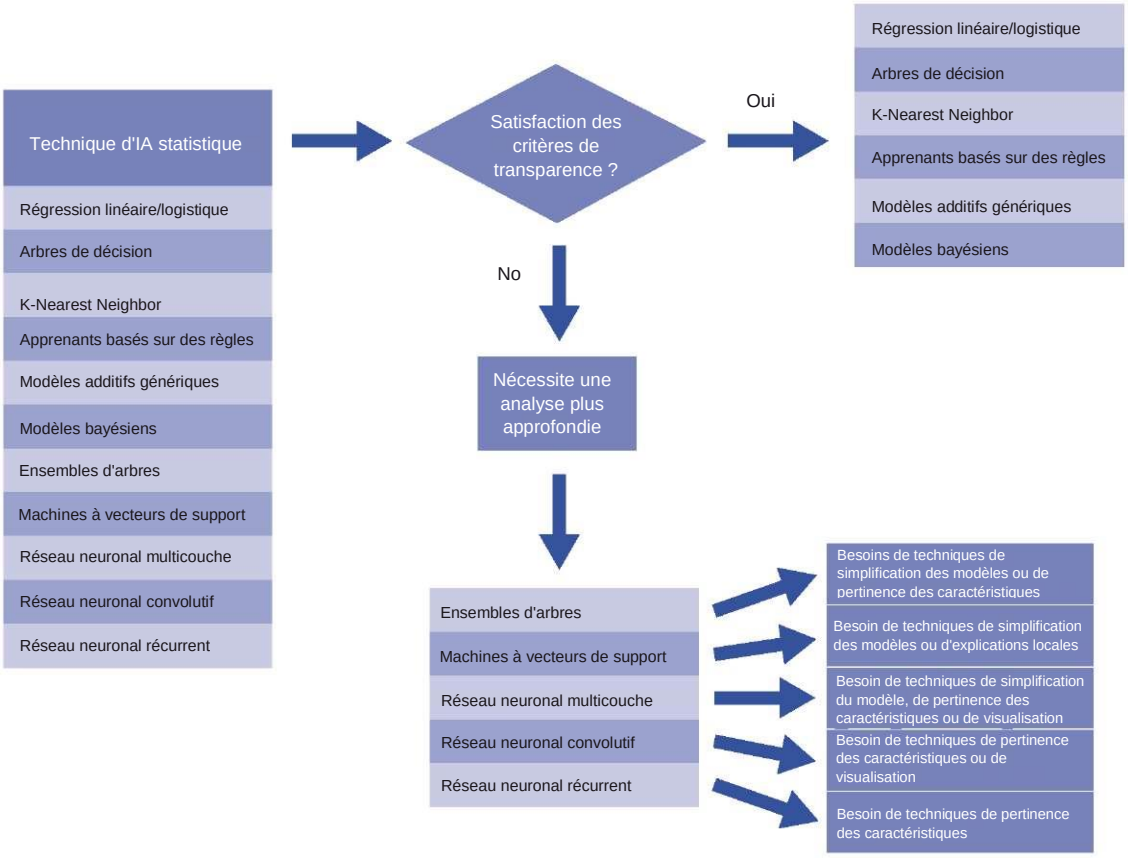
- **Simulabilité** - Le modèle peut être raisonné par un humain. La simulabilité signifie que l'algorithme peut être présenté dans des formats visuels et textuels.
- **Décomposabilité** - Chaque partie de l'algorithme, de l'entrée des données au calcul, peut être expliquée. La décomposabilité signifie que toutes les parties de l'algorithme sont compréhensibles par les humains sans avoir besoin d'outils supplémentaires.
- **Transparence algorithmique -** La façon dont l'algorithme produit le résultat peut être comprise par un utilisateur. La transparence algorithmique signifie que l'algorithme est entièrement explicable mathématiquement.

Les principaux algorithmes d'IA statistique ne sont pas tous transparents.
Par conséquent, un travail supplémentaire est nécessaire pour parvenir à l'explicabilité des techniques les plus complexes, comme le montre la **figure 9.**[18]

Les techniques de visualisation référencées dans la **figure 9** ne sont pas nécessairement associées aux techniques modernes de visualisation de la business intelligence. Au contraire, elles sont généralement associées à des modèles d'IA de type boîte noire, tels que l'analyse de sensibilité basée sur les données, l'analyse de sensibilité Monte Carlo et l'analyse de sensibilité par grappes.[19], [20]

FIGURE 9

Explicabilité des principaux algorithmes d'IA statistique



Explicabilité de l'IA

L'IA responsable a été définie comme une IA qui est robuste, explicable, éthique et efficace.[21] Le principe d'explicabilité est un élément clé du RGPD, qui énonce le droit des personnes à obtenir une explication pour toute forme de traitement des données les concernant. [22] La discipline de recherche de l'IA explicable (XAI) vise à contrer l'idée que le traitement de l'IA est une boîte noire et que peu de personnes, voire aucune, comprennent comment la boîte noire produit son résultat.

L'explicabilité est probablement l'un des principaux obstacles à une adoption plus large de l'IA. L'une des raisons pourrait être l'écart entre la recherche sur l'IA et les mises en œuvre technologiques associées. Les industries et les administrations strictement réglementées des secteurs privé et public (par exemple, la banque, la finance, les titres financiers et la santé) « ont traditionnellement pris du retard dans la transformation numérique de leurs processus » en raison de leur réticence à mettre en œuvre des techniques qui pourraient mettre leurs actifs en danger.[23]

La figure 10 présente les objectifs de l'IA explicable et les types de parties prenantes que ces objectifs pourraient intéresser.[24]

La causalité, le caractère informatif, la confiance, l'équité et le respect de la vie privée sont considérés comme les éléments les plus intéressants pour les régulateurs. Cependant, cette liste n'est pas exhaustive ; il pourrait y avoir d'autres parties prenantes intéressées par les objectifs généraux de l'IA explicable, comme les parties prenantes identifiées pour la vérifiabilité de l'IA discutée.

Dans certaines circonstances, plus l'algorithme d'IA est interprétable, plus la précision du modèle est faible. [25] Par exemple, l'IA symbolique est hautement interprétable, mais il est probablement impossible d'atteindre l'échelle nécessaire pour créer des règles pour chaque situation possible. Les résultats peuvent être moins précis que ceux obtenus avec les algorithmes d'apprentissage profond de l'IA statistique, par exemple, qui sont difficiles à interpréter mais potentiellement plus précis. Cela ne veut pas dire que les modèles plus complexes sont toujours plus précis. Le fait est qu'en général, il existe un compromis entre l'interprétabilité des algorithmes d'IA et les performances.[26]

L'explicabilité de l'IA ne se limite pas aux cas dans lesquels le système d'IA fonctionne comme prévu. Dans les cas où elle ne fonctionne pas comme prévu, l'évaluation de l'IA **(figure 8)** est particulièrement importante, car les performances de la technologie doivent être contrôlées en permanence pour garantir son efficacité et sa précision.

FIGURE 10

Les différents intérêts des parties prenantes dans les objectifs de l'IA explicable

Objectif de l'IA explicable	Parties prenantes						
	Experts du domaine	Scientifiques des données	Développeurs	Propriétaires de produits	Utilisateurs	Direction et le Conseil	Régulateurs
Fiabilité	X				X		
Causalité	X				X	X	X
Transférabilité	X	X					
Caractère informatif	X	X	X		X	X	X
Confiance	X		X			X	X
Impartialité	X		X			X	X
Accessibilité				X	X	X	
Interactivité	X				X		
Sensibilisation à la protection de la vie privée					X		X

Conclusion

L'IA statistique est nettement plus difficile à contrôler et à expliquer que l'IA symbolique. En effet, les règles de l'IA symbolique sont faciles à suivre pour les humains, contrairement aux algorithmes de l'IA statistique, qui pourraient n'être compréhensibles que par des scientifiques spécialisés dans les données. En outre, l'un des défis propres à l'IA statistique est que des données sont nécessaires pour entraîner les algorithmes.

Étant donné que les données d'entrée de l'IA statistique ne sont jamais parfaites (données sales), il existe de nombreux résultats aberrants potentiels qui pourraient compromettre l'efficacité d'un outil d'IA. Une tragédie basée sur un algorithme pourrait faire en sorte que l'ensemble de l'initiative d'IA soit perçu négativement par le marché, la solution serait pourrait alors être considérée comme un succès technique mais un échec commercial en raison de limites technologiques qui n'ont jamais été explicitées à la communauté des utilisateurs. Il est également à craindre que des risques juridiques et financiers ne soient réalisés en raison de l'absence de contrôle et de gestion de ces limitations.

Cependant, les activités de vérifiabilité et d'explicabilité ne sont pas requises pour toute l'IA statistique.

Compte tenu des efforts à fournir pour rendre l'IA vérifiable et explicable, il est préférable de réserver ces activités à l'IA à risque élevé ou limité. Considérer l'IA dans le contexte du risque qu'elle représente pour les communautés qu'elle sert est essentiellement une approche de la surveillance de l'IA basée sur le risque, que la vérifiabilité contribue à soutenir.

Les exigences minimales de vérifiabilité pour les sept domaines principaux de la **figure 3** comprennent un support documentaire détaillé pour le risque identifié (avec les contrôles), les actions entreprises, les problèmes et les questions traités, et les décisions qui ont abouti à l'approche spécifique adoptée dans chaque cas. Il convient de noter que la vérifiabilité ne concerne pas seulement les algorithmes d'IA, mais aussi leurs entrées, leurs sorties et leurs conséquences.

Les activités qui permettent de réaliser une IA explicable ne sont pas une panacée pour le problème troublant de la boîte noire de l'IA. Les préjugés et les failles de sécurité (y compris les injections de données) peuvent avoir un impact sur les performances de l'algorithme d'IA et brouiller les résultats, par exemple. Pourtant, certains partisans de l'IA ont suggéré qu'il n'est pas nécessaire de comprendre entièrement le fonctionnement de la boîte noire pour pouvoir profiter des avantages de la technologie, surtout si l'efficacité globale du système d'IA peut être démontrée à l'aide de mécanismes alternatifs.

Bien que les activités réglementaires semblent être à l'origine des exigences de vérifiabilité et d'explicabilité de l'IA dans de nombreux cas, il existe d'autres considérations importantes. Du point de vue commercial,

il est utile de poursuivre des activités qui contribuent à garantir que l'IA statistique fait ce qu'elle est censée faire dans une variété de scénarios. Cette approche permet de réduire les risques pour la communauté des utilisateurs ou la communauté sur laquelle l'IA a un impact, tout en protégeant l'organisation propriétaire de la technologie.

Même si le raisonnement est équivalent à « Si vous ne pouvez pas l'expliquer, alors vous ne le comprenez pas », on ne peut pas se tromper en recherchant la vérifiabilité et l'explicabilité d'un système d'IA. L'objectif est de veiller à ce que l'humanité ne subisse pas l'impact négatif d'une technologie qui a tant de potentiel pour faire le bien.

Bibliographie

1 Commission européenne, « Une approche européenne de l'intelligence artificielle », 23 février 2022, https://digital-strategy.ec.europa.eu/en/policies/ european-approach-artificial-intelligence

2 Commission européenne, *Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées en matière d'intelligence artificielle (loi sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union,* Belgique, 21 avril 2021, *https://eur-lex.europa.eu/resource. html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF*

3 Commission européenne, « Proposition de cadre réglementaire sur l'intelligence artificielle », 28 février 2022, *https://digital-strategy. ec.europa.eu/en/policies/regulatory-framework-ai*

4 Commission européenne, *Annexes à la proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées en matière d'intelligence artificielle (loi sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union,* Belgique, 21 avril 2021, *https://eur-lex.europa.eu/resource. html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_2&format=PDF*

5 Friedrich, S.; G. Antes; S. Behr; *et al.;* « Is There a Role for Statistics in Artificial Intelligence? » Advances in Data Analysis and Classification, 2021, *https://link.springer. com/article/10.1007/ s11634-021-00455-6*

6 Pearce, G.; M. Kotopski; "Algorithms and the Enterprise Governance of AI », *Journal de l'ISACA®,* vol. 4, 2021, *https://www.isaca.org/archives*

7 Zoldi, S.; « It's 2021. Do You Know What Your AI Is Doing? » FICO Blog, 25 mai 2021, *https://www. fico.com/blogs/its-2021-do-you- know-what-your-ai-doing*

8 Pearce, G.; « Digital Transformation? Boards Are Not Ready for It! » *Journal de l'ISACA,* vol. 5, 2018, *https://www.isaca.org/archives*

> « L'objectif est de s'assurer que l'humanité n'est pas impactée négativement par une technologie qui a tant de potentiel pour faire le bien. »

9 Zoldi, S.; « Establish AI Governance, Not Best Intentions, to Keep Companies Honest », *InformationWeek,* 30 novembre 2020, *https://www.informationweek.com/ai-or-machine-learning/establish-ai-governance-not-best- intentions-to-keep-companies-honest*

10 *Op cit* Pearce et Kotopski

11 Scanlon, L. ; « Auditability of AI Vital for Financial Services », Pinsent Masons, 15 février 2021, *https://www.pinsentmasons.com/out-law/analysis/auditability-of-ai-financial-services*analysis/auditability-of-ai-financial-services

12 Zoldi, S. ; « Beyond Responsible AI : Eight Steps to Auditable Artificial Intelligence », FICO Blog, 22 juin 2021, *https://www.fico.com/blogs/ beyond-responsible-ai-8-steps-auditable- artificial-intelligence*

13 *Op cit* Scanlon

14 Institutions supérieures de contrôle de Finlande, d'Allemagne, des Pays-Bas, de Norvège et du Royaume-Uni, « Auditing Machine Learning Algorithms », 24 novembre 2020, *https://www.auditingalgorithms.net/AuditabilityChecklist.html*

15 *Ibid.*

16 Hackernoon, « What Is Auditability for AI Systems? » 5 juillet 2021, *https://hackernoon.com/ what-is-auditability-for-ai-systems-wnz3714*

17 Barredo-Arrieta, A.; N. Dfaz-Rodnguez; J. Del Sera; *et al.;* « Explainable Artificial Intelligence (XAI) : Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI », *Information Fusion,* 9 janvier 2020, *https://www. sciencedirect.com/science/article/abs/pii/S1566253519308103*

18 *Ibid.*

Ibid.

19 Cortez, P.; M. Embrechts; « Opening Black Box

20 Data Mining Models Using Sensitivity Analysis », Institute of Electrical and Electronics Engineers (IEEE), 31 mars 2011, *https://core.ac.uk/download/pdf/55616214.pdf*

21 *Op cit* Zoldi, juin 2021

22 Koerner, K.; « Privacy and Responsible AI », The International Association of Privacy Professionals (IAPP), 11 janvier 2022, *https://iapp.org/news/a/privacy-and-responsible-ai/*

23 *Op cit* Barredo-Arrieta *et al.*

24 *Ibid.*

25 *Ibid.*

26 *Ibid.*