

L'intelligence artificielle et les algorithmes d'IA

4. L'auditeur interne et les algorithmes d'IA

2 bonnes raisons pour les auditer

Communauté
Intelligence
Artificielle



Production de
JM. Pluzanski

Risques IA

- Biais
- Opacité des modèles
- Corrélation erronée des données

Les devoirs de l'audit interne (IIA)

Recommandations classiques

- inclure l'IA dans son évaluation des risques
- fournir une assurance sur la gestion des risques liés à la fiabilité des algorithmes sous-jacents et des données sur lesquelles reposent les algorithmes
- évaluer si les activités d'IA réalisent leurs objectifs
- s'assurer que les questions éthiques et morales qui peuvent entourer l'utilisation de l'IA par l'organisation sont traitées

IA - Considérations pour la profession d'audit interne

Pourquoi auditer les algorithmes d'IA ?

A l'heure où les algorithmes d'IA occupent le devant de la scène, avec en corollaire la nécessité de les auditer compte tenu des risques annoncés en termes d'éthique, de biais et de boîtes noires, il est important pour les auditeurs internes de trouver leurs propres sources de motivation et les bonnes raisons pour s'inscrire maintenant dans cette démarche.

Après tout, pourquoi les entreprises novatrices qui utilisent des algorithmes d'IA dans leurs activités seraient-elles exposées à des risques supérieurs à celles qui s'appuient davantage sur « l'homme » susceptible pourtant de préjugés et de limites ?

Quel sera l'impact de l'IA sur la profession d'audit interne ? Quelles seront les missions des auditeurs internes face aux algorithmes d'IA ?

Des recommandations de l'IIA qui donnent le cap

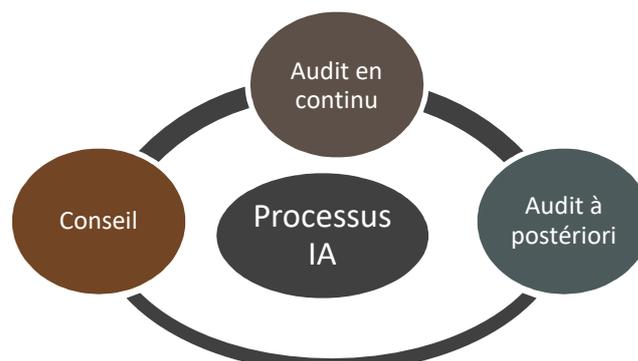
Les normes internationales de l'IIA, complétées par les recommandations de la publication « IA - considérations pour la profession d'audit interne », donnent aux auditeurs internes le cap à suivre.

L'audit de la fiabilité des algorithmes d'IA et de leurs données ne constituera pas le seul fondement de l'assurance délivrée par les auditeurs internes. Elle devra être augmentée par leurs participations actives en tant que conseils aux projets d'IA, et ce dès leurs débuts.

« Participer activement aux projets d'IA dès leurs débuts, en fournissant des conseils et des points de vue qui contribuent à la réussite de leur mise en œuvre »

Cette approche novatrice illustre l'impact disruptif de l'IA sur l'Audit interne et le modèles des 3 lignes de maîtrise.

L'opacité des algorithmes peut effectivement nécessiter que les auditeurs internes interviennent comme **conseils** à la création des processus d'IA. IBM a aussi annoncé récemment « l'audit en temps réel de l'IA » avec des outils ouvrant la « boîte noire », et capable d'expliquer les décisions des algorithmes.



- **Comprendre et analyser les processus d'IA dès leur conception pour ne pas devenir des auditeurs aveugles**

L'IA, un changement de paradigme pour les auditeurs internes

Les auditeurs internes ont un atout majeur dans leurs missions ; ils sont sur le terrain à la fois au contact des managers qui ont conçus les procédures et des collaborateurs qui doivent les appliquer. Les entretiens, menés sur cette double échelle, leur permettent pendant la mission d'audit d'apprécier la pertinence des procédures, et d'observer si elles sont comprises et partagées par les collaborateurs.

Avec le développement de l'IA, les auditeurs internes seront confrontés à un changement de paradigme. Les algorithmes visent des processus qui gèrent à grande vitesse un volume de données très important sur la base de modèles mathématiques complexes. Ils sont souvent conçus et développés à l'extérieur des entreprises par des start-ups en nouvelles technologies ou des entreprises leaders en informatique.

L'IA réduira le nombre de collaborateurs intervenant directement dans les processus, et la compréhension des modèles mathématiques les alimentant sera limitée à un cercle restreint. Il sera alors très difficile aux auditeurs internes de comprendre, lors de missions ponctuelles, ces processus d'IA basés sur l'apprentissage automatique, et de délivrer une assurance robuste.

Les auditeurs doivent être invités à la table de discussion sur les projets d'IA

Sous peine d'être « aveugles », ils doivent comprendre le fonctionnement des algorithmes d'IA en « **participant activement aux projets d'intelligence artificielle dès leurs débuts en fournissant des conseils et des points de vue** ». Ils pourront participer également aux tests d'essais avant la mise en place du système d'IA.

Les missions ponctuelles d'audit réalisées à posteriori relèveront davantage de la surveillance de l'utilisation de l'IA dans un cadre déjà analysé ; objectifs du processus d'IA, qualités des données utilisées, tests d'essais ...

Il faudra sans doute convaincre les dirigeants et les développeurs d'IA pour que les auditeurs internes aient une place à la table de discussion sur les projets d'IA. La question de leur légitimité à intervenir en amont avant même la mise en place des processus d'IA sera posée. Aussi cette volonté des auditeurs internes doit être relayée auprès des dirigeants des entreprises et des organes de contrôles.

- L'IA : un changement de paradigme pour les auditeurs internes
- La maîtrise des algorithmes d'IA nécessite une intervention des auditeurs au stade des projets
- Les auditeurs doivent être invités à la table de discussion sur les projets d'IA

- **Développer le leadership et la valeur ajoutée des auditeurs internes au sein des entreprises 4.0**

Cathy O'Neil

- Aucune excuse à ne pas auditer les algorithmes

CNIL

- Corps public d'experts des algorithmes pour les contrôler
- Homologation d'entreprises d'audit privées sur la base d'un référentiel

Rapport Villani

- Fonction assermentée d'audit des algorithmes

AI Now Institute at NY University 2017

- Normes strictes pour l'audit des systèmes d'IA « in the wild »

Etude pôle Fintech innovation ACPR

- Audit des algorithmes d'IA par le contrôle interne (permanent & périodique) ou par le superviseur

Un consensus pour auditer les algorithmes

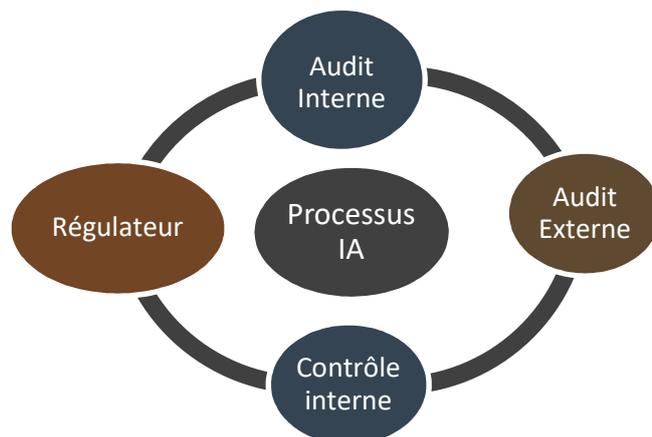
L'utilisation d'algorithmes d'IA sans réflexion éthique et sans contrôles peut générer des risques majeurs à très grande échelle. La nécessité d'auditer les algorithmes d'IA fait à ce jour consensus pour les différents acteurs du numérique et les organes de contrôles (Cf. annexe : Etudes de place).

Cathy O'Neil, ex-chercheuse en mathématiques aux USA, puis analyste dans un fonds d'investissement, a lancé la première l'alerte sur l'effet néfaste des algorithmes dans son livre *Weapons of Maths Destruction*, paru en 2016 aux États-Unis puis publié en 2018 en France sous le titre « Algorithmes, la bombe à retardement ». La question de réglementer les usages de l'IA dans le secteur financier a été posée récemment par l'Autorité de contrôle prudentiel et de résolution (ACPR) dans son étude de décembre 2018 « Intelligence artificielle : enjeux pour le secteur financier »

L'audit interne, une solution envisagée mais parmi d'autres

Les solutions envisagées dans les différentes études de place sont variées et encore au stade de la réflexion ; de la création d'une fonction assermentée d'audit des algorithmes, ou d'un corps public d'experts, jusqu'à l'audit des algorithmes par des entreprises d'audit privées homologuées. L'ACPR évoque l'intervention des contrôleurs et auditeurs internes.

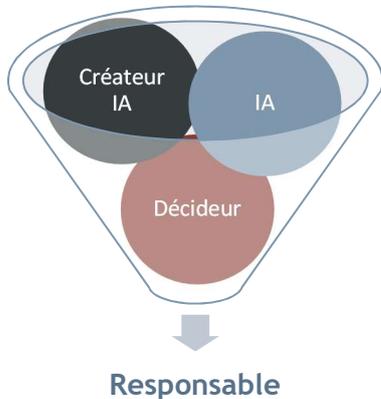
Les entreprises qui développent l'IA estiment quant à elles qu'elles peuvent s'auto-réguler. Aussi elles prennent les devants, désireuses que l'innovation ne soit pas freinée par l'ajout de nouvelles réglementations. Leurs chartes d'éthiques sur l'IA récemment publiées, sont très convergentes et font échos aux préoccupations soulevées en termes de risques (Cf. annexe : chartes de Google et de SAP). Certaines veulent aller plus loin. IBM par exemple a développé des outils pour ouvrir la « boîte noire » de l'IA. Ils doivent permettre à leurs clients de comprendre les décisions de l'IA, et de détecter les biais dans les algorithmes, « **comme un audit en temps réel de l'IA** » (Cf. annexe IBM Watson).



« Réguler une technologie naissante est difficile, mais je pense que les entreprises devraient s'auto-réguler »

Sundar Pichai - PDG de Google

• Développer le leadership et la valeur ajoutée des auditeurs internes au sein des entreprises 4.0



Quelle sera la place de l'auditeur interne dans les entreprises 4.0 ?

L'Audit interne doit prendre sa place dans la chaîne de valeur de l'IA pour développer son leadership et sa valeur ajoutée. Il apportera la confiance nécessaire aux dirigeants des entreprises 4.0 pour que les résultats de l'IA aident à la prise de décisions. Le NESTA recommande par exemple « qu'une **personne de la direction** doit être tenue **responsable** de toute mesure prise à la suite d'une décision algorithmique » (Cf. annexe : Etudes de place).

C'est une des raisons pour laquelle **l'audit des algorithmes s'impose aux auditeurs internes**. L'enjeu sera de taille car les entreprises 4.0 seront conduites « à grande vitesse », sur la base de l'interaction entre l'homme et l'IA. Ils devront assumer cette mission, en se formant aux nouvelles technologies, pour objectiver leurs travaux et ne pas se subordonner aux spécialistes en IA.

Des entreprises 4.0 à grande vitesse

Tout ira de plus en plus vite. Les algorithmes d'IA accéléreront la vitesse des processus et les prises de décisions. La temporalité entre les différentes phases d'exécution, de contrôles, de décisions, qui permettaient à l'homme de « corriger le tir » sera diminuée. Une fois mise en production, l'IA ne doutera jamais, car elle n'a pas de conscience. Aussi l'assurance délivrée par les auditeurs internes sur la fiabilité de l'IA deviendra essentielle.

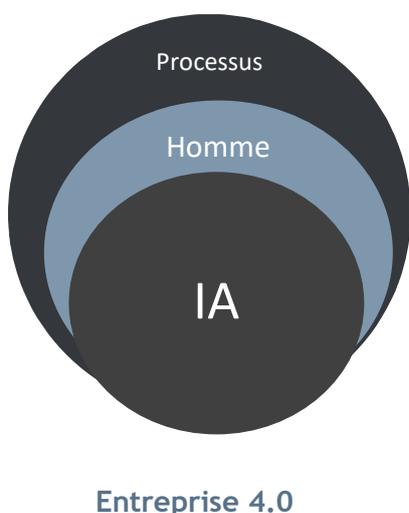
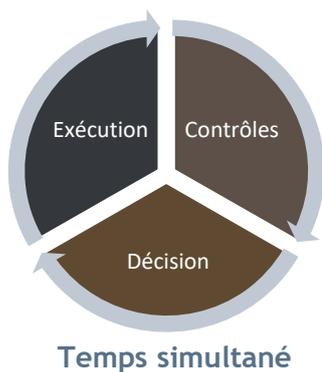
Une interaction entre l'homme et l'IA qui reste à définir

Deux facteurs principaux sont à prendre en compte ; la **part relative de l'homme et de l'IA** dans l'exécution des processus jusqu'à la prise de décisions, et la proportion de **travaux à forte valeur ajoutée** inclus dans les processus d'IA.

L'IA intégrera des tâches complexes en termes de calculs, de corrélations sur des données à grande échelle pour des aides à la décision. Elle inclura aussi au cœur des activités des contrôles sur les risques jusque-là assurés par la 2^e ligne de maîtrise (contrôle interne ...). Si on ajoute les tâches à faible valeur ajoutée qui peuvent être automatisées, la part de l'homme devrait diminuer dans l'exécution des processus.

La conception de l'IA par l'homme ne suffira pour éliminer les risques liés une IA sans conscience dans la vie réelle. Le risque spécifique de dérive lié à l'apprentissage automatique est aussi à prendre en compte (source ACPR).

Plus l'IA se substituera à l'homme dans les processus, plus les entreprises 4.0 auront besoin de l'Audit Interne pour veiller au maintien d'une conscience humaine indispensable.



• Développer le leadership et la valeur ajoutée des auditeurs internes au sein des entreprises 4.0

- Agilité
 - conseil et contrôle
 - compréhension de l'humain
- Ecoute
 - détection des signaux faibles de l'IA
 - dialogue avec les utilisateurs
- Objectivité
 - avoir un état d'esprit non biaisé s'inscrit dans l'ADN de l'auditeur interne

Les auditeurs internes : des acteurs de confiance et légitimes pour auditer les algorithmes d'IA

Les entreprises 4.0 auront besoin de « gardiens » de l'IA pour s'assurer qu'elle est maîtrisée. Les auditeurs internes doivent en assumer le leadership compte tenu de leurs compétences en termes d'agilité, d'écoute, d'objectivité et indépendance.

Agilité

L'Audit Interne est une fonction agile qui a évolué en fonction des besoins des entreprises. Les auditeurs internes sont devenus des conseillers de confiance, au-delà du strict périmètre de leurs missions d'audit. Cette double compétence en termes de conseil et de contrôle, qui repose sur leurs connaissances précises des activités, des processus et des collaborateurs est nécessaire aux entreprises pour développer leurs projets d'IA.

Ils ont aussi acquis, au cours de leurs missions, une grande agilité dans la compréhension du fonctionnement humain et des risques associés. L'audit des algorithmes d'IA créés par l'homme et au profit de l'homme, s'inscrit naturellement dans leur champ d'intervention.

Ecoute

Quelle écoute sera accordée aux collaborateurs ou clients qui douteront des résultats de l'IA ? Voudront-ils s'exprimer au risque qu'un « intervenant humain engagera davantage sa responsabilité à contredire le résultat d'un algorithme qu'à le valider » ? Auront-ils la connaissance suffisante des algorithmes pour objectiver leurs doutes ?

Les auditeurs internes, immergés dans la culture et les activités de l'entreprise, seront en capacité de détecter les signaux faibles des risques de l'IA (biais, erreurs de logiques...). Leurs participations en amont aux projets d'IA permettront un dialogue efficace avec les utilisateurs des algorithmes.

Indépendance et objectivité

« Avoir un état d'esprit non biaisé » s'inscrit dans l'ADN de l'auditeur interne. Son indépendance caractérisée par l'accès direct à la Direction générale et au Conseil lui permet d'exercer sans biais ses responsabilités en termes d'IA. Son objectivité fait qu'il ne subordonnera pas son jugement professionnel à celui des sociétés développeuses d'IA.

Charte éthique IA



- Être socialement bénéfique
- Éviter de créer ou de renforcer les préjugés injustes
- Être conçue et testée pour la sécurité
 - Tests de l'IA dans des environnements contraints
 - Surveillance du fonctionnement de l'IA après déploiement
- Faire face à ses obligations
 - Conception de systèmes d'IA qui offrent des possibilités appropriées de rétroaction, d'explications pertinentes et d'appel
 - Direction et contrôle humain des systèmes d'IA
- Respecter la vie privée
- Mise à disposition de l'IA pour des utilisations conformes à ces principes
- Maintenir un niveau élevé d'excellence scientifique

Charte éthique IA



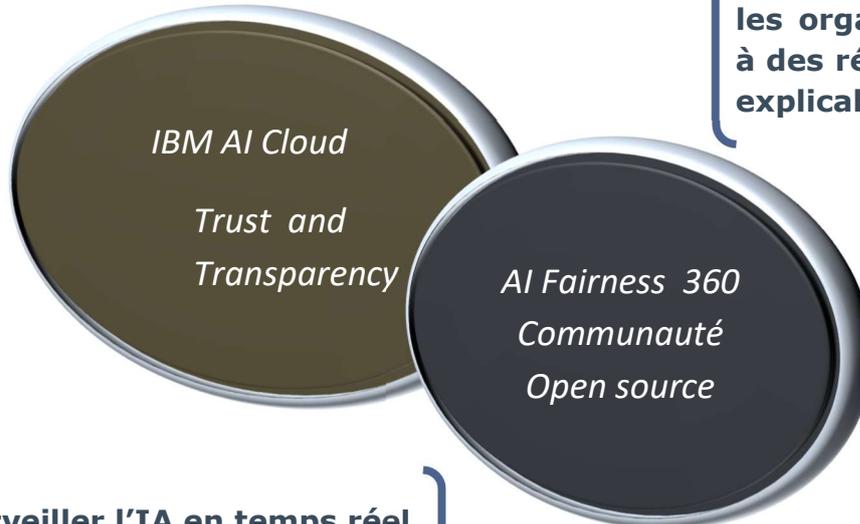
- Engagement pour la société au sens large
- Permettre aux entreprises de faire des affaires sans biais
 - Qualité des données
 - Diversité et interdisciplinarité des équipes
 - Etudes de nouvelles méthodes techniques pour atténuer les biais
- Respect de normes de qualité et de sécurité
 - Tests approfondis de l'IA dans le cadre de scénarios réels
 - Travail en étroite collaboration avec nos clients et utilisateurs
- Recherche de la transparence et de l'intégrité
 - Communication aux clients des capacités, objectifs et limites de l'IA
 - Fourniture des moyens de surveillance et de contrôle de l'IA par les clients et les utilisateurs
- Protection des données et de la vie privée
- Conception pour augmenter les talents des utilisateurs
 - Etroite collaboration avec les utilisateurs dans un environnement collaboratif, multidisciplinaire et démographiquement diversifié



Les outils

Permettre de comprendre, vérifier et expliquer la prise de décisions de l'IA

Accélérer et optimiser l'utilisation de l'IA dans les organisations grâce à des résultats fiables et explicables



Surveiller l'IA en temps réel

- exactitude
- performance
- équité

Détecter et contrer les biais nuisibles

Les enjeux

IBM Institute
for Business
Value AI 2018
report
(5 000 cadres
supérieurs)

82 % des entreprises envisagent d'adopter l'IA avec pour objectif l'accroissement du chiffre d'affaires

63 % n'ont pas la compétence
60 % craignent des problèmes de responsabilité

Domaines à plus grande valeur ajoutée de l'IA

Informatique
Sécurité de l'information
Innovation
Service à la clientèle
Gestion des risques

IBM Watson : outils de contrôle des algorithmes

Confiance

- Clients
- Auditeurs
- Conformité
- Régulateurs

Audit en temps réel de l'IA

"L'utilisateur pourra comprendre en temps réel comment le modèle d'IA a abouti à une décision.

Il pourra accéder, dans des termes clairs, aux éléments de recommandation utilisés par l'IA, et comprendre sur quels faits et données ils sont basés.

C'est comme un audit en temps réel de l'IA, une traçabilité qui permet de donner du sens aux décisions sur des éléments dignes de confiance et à valeur ajoutée, la décision pouvant être refusée si le processus est jugé insatisfaisant."

"Dans de nombreuses industries, l'examen réglementaire constitue un obstacle important à l'adoption de l'IA. Même si une entreprise est convaincue que ses modèles sont équitables et qu'elle peut faire confiance aux résultats, les organismes de réglementation exigent souvent une approche plus rigoureuse.

Pour cette raison, il est essentiel de s'assurer que la contribution de l'IA à toute décision est entièrement explicable en gardant une trace complète de la lignée de tous les modèles, données, entrées et sorties de toute application alimentée par l'IA ..."

"Si l'intelligence artificielle est une boîte noire qui ne fait que recueillir des données et produire des résultats, il n'y a pas de moyen facile pour l'entreprise de juger si ces résultats sont un bon guide pour la prise de décision ou non. De même, l'entreprise ne sera pas en mesure d'expliquer les résultats aux clients, aux auditeurs ou aux équipes de conformité. "

" Presque tous les modèles d'IA, aussi soigneusement conçus soient-ils, sont susceptibles de présenter un certain degré de biais.

La qualité d'un modèle dépend des données sur lesquelles il est formé, et comme les ensembles de données de formation ne peuvent jamais être représentatifs à 100 % des données du monde réel, il y a toujours un risque qu'un modèle nouvellement formé ne fonctionne pas bien dans la production.

De plus, comme la plupart des domaines de données évoluent continuellement, l'exactitude des modèles tend à dériver avec le temps..."

"Les biais sont inhérents à l'IA dans la mesure où elle est programmée par des hommes et des femmes. Notre système permet de détecter le biais au moment même où il apparaît.

Par exemple, lorsqu'il suggère une décision sur les bases des jeux de données qu'il a à sa disposition, par exemple refuser une police d'assurance à une personne de 22 ans, il va aussi informer l'assureur que cette police a été approuvée à 91% pour les clients entre 31 ans et 55 ans, mais seulement à 52% pour les 18-24 ans.

Cette information devient enfin visible, en temps réel, ce qui permet au client de la prendre en compte"

" Une entreprise s'expose à un risque important s'il délègue des responsabilités à une IA qui ne correspond pas tout à fait aux attentes de l'entreprise.

Par exemple, il pourrait en résulter un grave préjudice financier ou de réputation si un modèle discrimine injustement un groupe particulier de clients parce que ses données de formation ne représentaient pas un échantillon suffisamment important de cette population "

Équité

- Détection des biais en temps réel



NEW YORK UNIVERSITY

Recommandations Report 2017 AI Now Institute at New York University

- Non utilisation par les organismes publics (justice pénale, santé, aide sociale, éducation) d'algorithmes d'IA reposant sur des boîtes noires
- Réalisations d'essais rigoureux par les entreprises avant le lancement d'un système d'IA pour s'assurer qu'ils n'amplifient pas les biais et les erreurs dus aux données de formations, et aux algorithmes
- Surveillance par les entreprises de l'utilisation du système d'IA, après son lancement
- Élaborer des normes pour suivre la provenance, l'élaboration et l'utilisation des données de formation de l'IA tout au long de leur cycle de vie pour mieux comprendre et surveiller les questions de biais
- Etablissement nécessaire et urgent de normes strictes pour l'audit et la compréhension de l'utilisation des systèmes d'IA « in the wild »



ACPR : Enjeux de l'IA pour le secteur financier

- Conformité des objectifs des algorithmes d'IA aux principes de gouvernance posés par les réglementations
- Recours à des experts pour valider la pertinence des variables utilisées, éliminer celles qui sont inutiles ou sources de biais potentiels
- Emploi d'un processus parallèle plus sûr et plus traditionnel sur une partie des données tests
- Utilisation d'un jeu de données étalon sur les algorithmes pour contrôler régulièrement à la fois la pertinence et l'aspect non-discriminatoire des algorithmes
- Développement d'outils qui évalueraient la dérive conceptuelle pour maîtriser le risque spécifique de l'apprentissage automatique
- Expliquer les mécanismes et les critères suivis par l'algorithme au cours de son processus d'analyse
- Tests à définir sur des jeux de données indépendants de ceux utilisés pour l'apprentissage des algorithmes
- Audit des algorithmes d'IA par le contrôle interne ou par le superviseur



NESTA - Fondation pour l'innovation britannique

10 recommandations pour encadrer la prise de décision algorithmique dans le secteur public

- Les citoyens doivent être informés lorsque leur traitement a été en partie ou entièrement décidé par un algorithme
- Tout algorithme utilisé par une organisation du secteur public devrait être accompagnée d'une description de ses fonctions, objectifs et impacts attendus, mis à disposition de ceux qui l'utilisent
- Les organisations du secteur public devraient publier les détails décrivant les données avec lesquels l'algorithme a été entraîné et les hypothèses utilisées à sa création ainsi qu'une évaluation des risques pour atténuer ses biais potentiels
- La liste de toutes les données d'entrées utilisées par l'algorithme pour prendre une décision devrait être publiée
- Les organisations du secteur public devraient s'engager à évaluer l'impact de leurs algorithmes et à publier les résultats de ces études
- Les algorithmes devraient être catégorisés sur une échelle de risque (de 1 à 5) pour distinguer leur impact sur les individus
- Tout algorithme doit être disponible en version de test, afin que les auditeurs puissent tester leur impact (et confirmer qu'il fait ce qu'il prétend notamment pour ceux qui ne peuvent pas être ouverts par défaut)
- Les organisations du secteur public qui souhaitent adopter des procédures de prise de décision algorithmique à haut risque devraient souscrire une assurance capable d'offrir des compensations aux personnes négativement touchées par une décision algorithmique erronée
- Une personne de la direction doit être tenue responsable de toute mesure prise à la suite d'une décision algorithmique
- Lorsque les services publics ont recours à des tiers pour créer ou exécuter des algorithmes en leur nom, ils devraient s'assurer que ces tiers respectent ces principes

Transparence

Risque
Et
Audit

Décision