

PERSPECTIVES INTERNATIONALES

Cadre de référence IIA pour l'audit de l'intelligence artificielle

Applications pratiques, Partie B

Édition spéciale



Comité consultatif

Nur Hayati Baharuddin, CIA,
CCSA, CFSA, CGAP, CRMA –
IIA–Malaisie

Lesedi Lesetedi, CIA, QIAL –
African Federation IIA

Hans Nieuwlands, CIA, CCSA,
CGAP – IIA–Pays-Bas

Karem Obeid, CIA, CCSA, CRMA
– IIA–Emirats Arabes Unis

Carolyn Saint, CIA, CRMA, CPA –
IIA – Amérique du Nord

Ana Cristina Zambrano Preciado,
CIA, CCSA, CRMA – IIA-Colombie

Numéros précédents

Pour accéder aux numéros
précédents de Perspectives
Internationales, visitez le site à
l'adresse suivante www.theiia.org/gpi.

Commentaires des lecteurs

N'hésitez pas à nous faire
parvenir vos questions et vos
commentaires à l'adresse
suivante :
globalperspectives@theiia.org

Table des matières

Introduction	1
Le Cadre de référence IIA pour l'audit de l'intelligence artificielle ...	1
Facteur humain	2
Éthique	2
Algorithmes défectueux	3
Donner du sens	3
Boîte noire	4
Gouvernance	5
Mesure de la performance	5
Conclusion	6

Introduction

Imaginez. Un robot – ou au moins son cerveau – comparissant devant un comité d'éthique ou d'enquête. Cela pourrait arriver. Selon [The Guardian](#), les universitaires avancent que, comme les robots commencent à investir les espaces publics et à travailler aux côtés des humains, le besoin d'adopter des mesures de sécurité se fait plus pressant. Les scientifiques plaident pour que les robots soient dotés d'une « boîte noire éthique » et conservent une trace de leurs décisions qui leur permette d'expliquer – oui, expliquer – leurs actions en cas d'accident.

Des professeurs de l'Université d'Oxford estiment que les entreprises spécialisées en robotique devraient suivre les exemples et les règles établis par le secteur de l'aviation notamment l'installation de boîtes noires et d'enregistreurs de conversation dans le poste de pilotage pour faciliter les enquêtes sur les crashes aériens et tirer tous les enseignements de ces tragédies dans le domaine de la sécurité. En étant installée sur un robot, une boîte noire éthique enregistrerait les décisions de ce dernier et leurs fondements, ses mouvements, ainsi que les informations provenant de capteurs, comme les caméras, les micros et les télémètres.

Ces actions s'inscrivent dans le sillage d'incidents récents, parmi lesquels on peut citer « Steve », un robot de surveillance K5 de [Knightscope](#), qui, après une chute dans les escaliers, s'est noyé dans une fontaine pendant l'exercice de ses fonctions, un autre robot K5 qui s'est retrouvé impliqué dans une altercation sur un parking avec un homme de 41 ans alors qu'il patrouillait les rues de Mountain View en Californie, ou encore une unité robotisée accusée d'avoir renversé un enfant de 16 mois dans un centre commercial à Stanford. Toutes les personnes concernées n'ont souffert que de blessures légères.

Il existe des centaines de rapports sur l'éthique de l'intelligence artificielle, mais la plupart sont légers et truffés de banalités sur l'importance de placer les humains au premier plan, écrit Scott Rosenberg, éditeur de *Backchannel*. Il se fonde sur des rapports publiés récemment par le *AI Now Institute* de l'Université de New York démontrant que le secteur des technologies tente de remodeler la société en s'inspirant de l'intelligence artificielle, sans aucune garantie sur la fiabilité et l'exactitude des résultats. Un rapport conclut que « jusqu'à ce jour, les efforts destinés à astreindre l'intelligence artificielle aux normes éthiques n'ont pas été concluants et que les nouveaux cadres éthiques pour l'intelligence artificielle doivent transcender la responsabilité individuelle afin d'amener les puissants intérêts industriels, gouvernementaux et militaires à rendre des comptes lorsqu'ils conçoivent et utilisent l'intelligence artificielle ». Selon l'auteur, les systèmes d'intelligence artificielle sont introduits dans des secteurs variés et vulnérables tels que la surveillance, l'éducation, les soins de santé et d'autres domaines où une « défaillance de l'algorithme pourrait détruire une vie ».

Le cadre de référence IIA pour l'audit de l'intelligence artificielle

Comme expliqué dans Intelligence artificielle – Considérations pour la profession d'audit interne, le rôle de l'audit interne est « d'aider une organisation à évaluer, comprendre et communiquer la mesure dans laquelle l'intelligence artificielle

Remarque : Il s'agit du troisième volet de la trilogie. Pour de plus amples informations, consultez les volets Intelligence artificielle – considérations pour la profession d'audit interne et Cadre de référence IIA pour l'audit de l'intelligence artificielle : Applications partie A



aura un effet (néгатif ou positif) sur la capacité de cette organisation à créer de la valeur à court, moyen, ou à long terme ».

Afin d'aider l'audit interne à remplir son rôle, les auditeurs internes peuvent s'inspirer du Cadre de référence IIA pour l'audit de l'intelligence artificielle lorsqu'ils fournissent des services de conseil ou d'assurance en lien avec l'intelligence artificielle, ou les deux, selon les besoins de l'organisation. Le Cadre de référence est constitué de trois composantes fondamentales, à savoir la stratégie relative à l'intelligence artificielle, la gouvernance et le facteur humain, et de sept sous-éléments : la cyber résilience, les compétences en intelligence artificielle, la qualité des données, l'architecture et l'infrastructure des données, la mesure de la performance, l'éthique et la boîte noire.

L'audit interne est tenu de considérer les nombreux objectifs de mission ou de contrôle ainsi que les activités ou procédures lors de la mise en œuvre du Cadre de référence et lors de la prestation de services de conseil, d'assurance, ou les deux, en lien avec les activités d'intelligence artificielle de l'organisation. Les objectifs et les activités ou procédures pertinents pour les volets stratégie (cyber résilience et compétences en intelligence artificielle) et gouvernance (architecture et infrastructure des données, qualité des données) du Cadre de référence ont été fournis dans le document « Cadre de référence IIA pour l'audit de l'intelligence artificielle : Applications pratiques Partie A ».

Cette dernière partie sera consacrée aux objectifs et activités ou procédures pertinents pour le facteur humain (éthique et boîte noire) et la gouvernance (mesure de la performance).

Facteur humain

La composante facteur humain, qui reprend les éléments éthique et boîte noire, aborde le risque d'une erreur humaine compromettant la capacité de l'intelligence artificielle à produire les résultats escomptés.

Éthique

Les algorithmes élaborés par des Hommes peuvent être affectés par des erreurs et des biais humains (intentionnels ou non), ce qui aura un impact sur leur performance. La composante facteur humain examine si :

- le risque que des biais humains involontaires interviennent dans la conception de l'intelligence artificielle est identifié et maîtrisé ;
- l'intelligence artificielle a été efficacement testée pour s'assurer que ses résultats sont conformes à l'objectif initial ;
- les technologies de l'intelligence artificielle peuvent être transparentes compte tenu de leur complexité ;
- le produit de l'intelligence artificielle est utilisé de manière légale, éthique et responsable.



Algorithmes défectueux

Selon un rapport récent de McKinsey & Company, les entreprises sont promptes à appliquer l'apprentissage automatique dans leur processus de prise de décision. Les programmes définissent des algorithmes complexes destinés à de vastes ensembles de données fréquemment actualisés. Cependant, s'il est négligé et non maîtrisé, le biais algorithmique comporte des risques dans la mesure où il compromet la raison d'être de l'apprentissage automatique (cf. *Controlling machine-learning algorithms and their biases*).

Par exemple, dans le cas de l'évaluation du risque de crédit, le client affichant un historique de prêts sans défaut de paiement est généralement qualifié de client à « faible risque ». Pourtant, le fait que les prêts hypothécaires de ce client ont été maintenus et soutenus grâce à des avantages fiscaux importants arrivant à échéance peut passer inaperçu. Un pic de défauts de paiement pourrait se profiler à l'horizon sans que le modèle de risque statistique de l'organisme de crédit ne le détecte. En ayant accès aux bonnes données et aux conseils d'experts, les modèles d'apprentissage automatique prédictifs pourraient déceler les tendances cachées dans les données et anticiper de telles hausses. En outre, en dehors des décisions prises au sein des entreprises, les biais algorithmiques peuvent créer des erreurs débouchant sur de réels problèmes et inquiétudes chez les citoyens. Par exemple, Google Photos, ou d'autres services similaires, permettent d'identifier des personnes, des objets et des lieux, mais peuvent totalement dysfonctionner, notamment lorsqu'une application de reconnaissance d'images souffre d'un bug à connotation raciste, ou lorsqu'un logiciel prédictif utilisé pour tenter d'évaluer les risques de récidive chez les prévenus s'avère biaisé.

Donner du sens

Il y a des limites à ce que les machines peuvent accomplir et les *Hommes* doivent être capables de donner du sens aux résultats de l'intelligence artificielle. Comme indiqué dans un rapport de McKinsey & Company, « *La quête de sens à l'ère de l'intelligence artificielle commence avec une appréciation de ce que les machines peuvent et ne peuvent pas faire. Il est possible, par exemple, qu'une machine pose certains types de diagnostics avec plus de précision qu'une personne. Mais il incombe aux infirmières, aux médecins et aux thérapeutes d'aider les patients à en comprendre les implications et à gérer les conséquences. Telle est la différence entre la connaissance et le sens* ». Selon le même rapport, des connaissances spécialisées comme le codage, l'analyse de données et la science des données sont indispensables à l'intelligence artificielle, au même titre que les aptitudes relationnelles, telles que la collaboration, l'empathie et la recherche de sens (cf. *McKinsey Quarterly* p. 56-61).

Les objectifs et les activités ou procédures pertinents identifiés par l'IIA ne comprennent pas un plan d'audit défini, mais constituent des exemples utiles pour identifier les objectifs de la mission ou du contrôle, et pour planifier et réaliser des missions d'audit dans le domaine de l'intelligence artificielle.

Ces dernières devraient respecter la Norme 2200 de l'IIA : Planification de la mission. Le plan d'audit et les objectifs et les procédures de la mission doivent toujours être adaptés pour répondre aux besoins de l'organisation.

Objectifs et activités ou procédures éthiques pertinents

Objectif(s) de la mission ou du contrôle	Activités ou procédures
Fournir l'assurance que les résultats des activités d'intelligence artificielle de l'organisation sont dépourvus de biais involontaires.	Revoir les résultats escomptés des activités d'intelligence artificielle (objectifs stratégiques) et les comparer avec les résultats réels. Si une différence est détectée, déterminer si un biais en est la cause.
L'organisation peut « donner du sens » aux résultats de l'intelligence artificielle.	Examiner les résultats de l'intelligence artificielle et leur interprétation.

Boîte noire

Selon le dictionnaire en ligne de *Merriam-Webster*, une boîte noire est « *un appareil électronique généralement complexe dont le mécanisme interne est habituellement caché ou mystérieux pour l'utilisateur; au sens large: tout ce qui a des fonctions ou mécanismes internes mystérieux ou inconnus* ». Plus les organisations progressent dans la mise en œuvre de technologies d'intelligence artificielle de Type III et IV, en utilisant des machines ou des plates-formes qui peuvent apprendre par elles-mêmes ou communiquer entre elles, moins le fonctionnement des algorithmes est transparent et compréhensible. Plus les activités d'intelligence artificielle des organisations gagneront en sophistication, plus le facteur boîte noire deviendra critique.

Objectifs et activités ou procédures pertinents concernant la boîte noire

Objectif(s) de la mission ou du contrôle	Activités ou procédures
Évaluer la compréhension de l'organisation des données de la « boîte noire » (les algorithmes sous-jacents, les fonctions internes ou les mécanismes qui rendent l'intelligence artificielle possible).	<p>Examiner les politiques de développement et de mise en œuvre de l'intelligence artificielle et vérifier que les données de la boîte noire ont été identifiées.</p> <p>S'entretenir avec les responsables des résultats de l'intelligence artificielle et vérifier qu'ils comprennent et sont capables d'expliquer les données de la boîte noire.</p>

Gouvernance

La gouvernance de l'intelligence artificielle désigne les structures, les processus et les procédures mis en place pour diriger, gérer et surveiller les activités liées à l'intelligence artificielle au sein de l'organisation. La structure et la formalisation de la gouvernance dépendront des caractéristiques spécifiques à l'organisation. La gouvernance de l'intelligence artificielle :

- établit l'obligation de rendre compte, la responsabilité et la supervision ;
- contribue à garantir que les responsables de l'intelligence artificielle disposent des compétences et de l'expertise nécessaires ;
- contribue à garantir que les activités, les décisions et les actions relatives à l'intelligence artificielle respectent les valeurs ainsi que les responsabilités éthiques, sociales et légales de l'organisation.

La Partie II de cette trilogie, « *Cadre de référence IIA pour l'audit de l'intelligence artificielle : Applications pratiques Partie A* », aborde la composante gouvernance, et plus particulièrement le devoir de rendre compte, la responsabilité et la supervision, la conformité réglementaire, l'architecture et l'infrastructure des données et la qualité des données. Ce document traite de la mesure de la performance.

Mesure de la performance

L'audit interne est bien placé pour devenir indispensable à l'organisation en lui permettant de mesurer la performance de ses initiatives en matière d'intelligence artificielle. Lors de la phase de planification, l'audit interne peut dispenser des conseils sur la manière d'établir les indicateurs qui fournissent à la direction et au Conseil des informations suffisantes, fiables, pertinentes et utiles. Cependant, l'audit interne ne doit pas être le responsable ou le propriétaire des indicateurs de performance de l'intelligence artificielle. Dans les organisations où l'intelligence artificielle a été mise en œuvre, l'audit interne doit fournir une assurance concernant les dispositifs de contrôle de la première ligne de maîtrise et de supervision de la deuxième ligne de maîtrise en lien avec l'intelligence artificielle. Il n'existe peut-être pas de meilleure façon pour l'audit interne de prouver sa compétence en la matière qu'en utilisant les technologies d'intelligence artificielle comme l'automatisation des processus robotisés pour auditer l'intelligence artificielle.



Objectifs et activités ou procédures pertinents pour mesurer la performance

Objectif(s) de la mission ou du contrôle	Activités ou procédures
Donner des conseils sur la manière d'établir les indicateurs de performance de l'intelligence artificielle.	Faciliter les séances de travail pour les personnes responsables d'établir les indicateurs de performance de l'intelligence artificielle. Communiquer sur l'importance et la signification des informations <i>suffisantes, fiables, pertinentes et utiles</i> .
Soumettre les vulnérabilités de l'intelligence artificielle à des tests de résistance.	Appliquer des techniques de tests de résistance utilisées par le secteur bancaire pour déterminer comment les activités d'intelligence artificielle réagissent dans des scénarios extrêmes.
Communiquer les résultats des missions en lien avec l'intelligence artificielle.	Communiquer les résultats des missions en lien avec l'intelligence artificielle conformément à : <ul style="list-style-type: none"> ■ La Norme de l'IIA 2400 : Communiquer les résultats ■ La Norme de l'IIA 2410 : Contenu de la communication ■ La Norme de l'IIA 2420 : Qualité de la communication ■ La Norme de l'IIA 2421 : Erreurs et omissions ■ la Norme de l'IIA 2430 : Utilisation de la mention « <i>conduite conformément aux Normes internationales pour la pratique professionnelle de l'audit interne</i> » ■ la Norme de l'IIA 2431 : Indication de non-conformité ■ la Norme de l'IIA 2440 : Diffusion des résultats
Fournir une assurance sur les dispositifs de contrôle de la première ligne de maîtrise et de supervision de la deuxième ligne de maîtrise en lien avec l'intelligence artificielle.	Adopter la robotique et d'autres formes d'intelligence artificielle pour réaliser des missions en lien avec l'intelligence artificielle.

Conclusion

La profession d'audit interne ne peut pas être à la traîne dans ce qui pourrait bien être le prochain cap de l'ère du numérique : l'intelligence artificielle. Pour se préparer, les auditeurs internes doivent comprendre les bases de l'intelligence artificielle, les rôles que l'audit interne peut et devrait jouer ainsi que les risques et les opportunités que représente l'intelligence artificielle.

Que les technologies et activités d'intelligence artificielle de l'organisation soient développées en interne, par le biais d'outils ou par un tiers, l'audit interne devrait être préparé à épauler le Conseil et la direction générale, à se coordonner avec les première et deuxième lignes de maîtrise et à fournir une assurance sur les dispositifs de management des risques, de gouvernance et de contrôle associés à l'intelligence artificielle. Pour relever ces défis, les auditeurs internes devraient s'inspirer du Cadre de référence IIA pour l'audit de l'intelligence artificielle.

À propos de l'IIA

Porte-parole mondial de la profession d'audit interne, l'*Institute of Internal Auditors* (IIA) est une autorité reconnue et un leader incontesté dans la formation et la formulation de normes, lignes directrices et certifications. Fondé en 1941, l'IIA compte actuellement quelque 190 000 membres dans plus de 170 pays et territoires.

Son siège se situe à Lake Mary (Floride) aux États-Unis. Plus d'informations sont disponibles sur le site www.globaliia.org ou www.theiia.org

Clause de non-responsabilité

Les opinions exprimées dans les Perspectives internationales ne sont pas nécessairement celles des auteurs ayant collaboré à l'élaboration du présent document ni celles des collaborateurs.

Copyright

Copyright © 2018 de l'Institute of Internal Auditors, Inc. Tous droits réservés.

