Risques induits par l'intelligence artificielle

Une approche d'aide à l'identification

Jacky Akoka[1], Isabelle Comyn-Wattiau[2]

1. Laboratoire CEDRIC-CNAM
2 Rue Conté, 75003 Paris, France
jacky.akoka@lecnam.net

2. ESSEC Business School
3 Avenue Bernard Hirsch, 95021 Cergy Cedex, France
isabelle.wattiau@essec.edu

RÉSUMÉ. L'intelligence artificielle est de plus en présente dans la vie des organisations. Elle offre beaucoup d'opportunités pour améliorer la performance de ces organisations. Son utilisation peut les exposer à des risques spécifiques, avec des conséquences potentiellement graves. Pour aider les organisations à repérer ces risques, cet article propose une approche d'aide à l'identification de ceux-ci. L'approche s'appuie sur un modèle conceptuel et trois typologies, respectivement des risques, des techniques d'intelligence artificielle et des processus métiers. Ces typologies sont croisées dans des matrices permettant d'identifier les menaces spécifiques. Deux matrices sont ainsi combinées pour déduire tous les risques potentiellement associés à l'usage de l'intelligence artificielle dans un processus métier. L'approche est illustrée à l'aide des processus métiers de l'assurance.

ABSTRACT. Artificial Intelligence is increasingly present in the life of organizations. It offers many opportunities to improve the performance of these organizations. However, its use can expose them to specific risks with potentially serious consequences. This paper proposes an approach helping organizations to identify these risks. This approach is based on a conceptual model and three typologies: risks, artificial intelligence techniques, and business processes. These typologies are cross-referenced in matrices to identify specific threats. Two matrices are combined to derive all the risks potentially associated with the use of artificial intelligence in a business process. The approach is illustrated using insurance business processes.

MOTS-CLES : intelligence artificielle, risque, donnée, approche méthodologique, typologie.

KEYWORDS: artificial intelligence, risk, data, methodological approach, typology.

1. Introduction

L'intelligence artificielle (IA) est présente dans tous les secteurs d'activité, qu'ils soient de nature industrielle ou des services. Selon Statista, le chiffre d'affaires du marché mondial des technologies de l'IA est estimé à 200 milliards de dollars en 2023 et devrait atteindre plus de 1800 milliards en 2030[1]. Ces technologies ne sont pas sans présenter un certain nombre de risques.

L'IA offre de nombreuses opportunités tant pour améliorer la productivité par l'automatisation des tâches que pour innover dans les produits et les services. L'IA contribue aussi à la prise de décision en fournissant les prédictions nécessaires à cette décision. Elle est présente dans le commerce électronique (personnalisation des achats, assistants intelligents, prévention des fraudes), dans l'éducation (contenus intelligents, assistants vocaux, apprentissage personnalisé), dans l'automobile (aide à la navigation, véhicule autonome, robotique). Elle pénètre aussi le domaine de la santé (détection des maladies, découverte de médicaments), les ressources humaines (repérage des talents), l'agriculture (analyse des données, robotisation). Elle offre aussi des applications créatives dans le domaine artistique ou dans le design. Enfin, sans pouvoir être exhaustif, mentionnons aussi la finance (aide à la gestion de patrimoine, détection des fraudes, analyse des risques).

Pour rendre possibles ces nouvelles applications, l'IA met en œuvre un certain nombre de techniques. Les techniques d'apprentissage automatique fondées sur l'exploitation des données permettent l'acquisition de connaissances. Elles peuvent être supervisées ou non, en milieu fermé ou en interaction avec l'environnement. Elles permettent de transférer une connaissance d'un domaine à l'autre, comme le raisonnement par analogie. La vision assistée par ordinateur reproduit la capacité de perception des images par l'homme. Le traitement de la langue naturelle permet de transférer à la machine les capacités humaines de communication. Plus récemment, l'IA générative a montré la capacité de l'ordinateur à générer des nouveaux contenus (textes, images, vidéos, etc.) par compilation de grands modèles d'informations pré-enregistrées.

Ces opportunités et les techniques qu'elles exigent ne sont pas sans générer leur lot de risques. Certains de ces derniers découlent de l'usage de données dont la qualité n'est ni totale ni connue. Par là-même, l'IA s'appuie sur une information incomplète ou erronée qui peut biaiser son raisonnement. D'autres risques sont de nature juridique dans la mesure où la transparence n'est pas toujours possible. Même transparent, un processus d'intelligence artificielle peut ne pas être conforme aux réglementations liées à l'usage des données tout au long de leur cycle de vie. C'est une révolution qui va aussi engendrer des mutations profondes sur le marché de l'emploi, laissant sur le côté celles et ceux qui voient leurs tâches effectuées plus facilement par des robots fiables, résilients et obéissants. L'IA requiert

[1] https://www.statista.com/topics/3104/artificial-intelligence-ai-worldwide/#topicOverview

l'accumulation de masses de données qu'il est difficile de gérer et contrôler, engendrant des craintes fondées quant au respect de la vie privée.

Parmi tous ces risques, il convient aussi de mentionner le risque à ne pas faire, laissant la concurrence prendre le dessus. C'est la raison pour laquelle toutes les organisations sont concernées par l'anticipation de ces risques. La question de recherche ciblée dans cet article est : comment peut-on identifier les risques liés à l'utilisation de l'IA dans un processus métier ? L'objet de cet article est de proposer une approche permettant de recenser et comparer les risques liés à l'usage de l'intelligence artificielle. L'approche doit être évolutive, généralisable à différentes activités et différents secteurs. Elle doit faciliter la tâche de compréhension du risque et de structuration de la réponse.

Le reste de cet article est organisé comme suit. La section 2 présente un état de l'art sur les techniques de l'IA et les risques associés. La section suivante décrit l'approche en déroulant chacune de ses étapes. Elle est instanciée au domaine de l'assurance. La dernière section est consacrée à la conclusion et aux recherches futures.

2. Revue de la littérature

Dans cette section, nous présentons et analysons la littérature relative aux typologies des techniques d'IA et celles relatives aux risques induits.

2.1. Sur les typologies de l'intelligence artificielle

De nombreuses typologies de l'IA sont présentées dans la littérature. Certaines sont fondées sur les capacités. Par exemple, ont été successivement définis les termes Intelligence Artificielle, puis IA Générale (qui effectue toutes les tâches intellectuelles dont l'humain est capable, inclus les tâches de créativité), puis, par réaction, l'IA étroite (restreinte à des tâches spécifiques, par exemple la reconnaissance faciale) et la super IA (qui peut surpasser l'intelligence humaine). Selon (Pereira *et al*., 2023), l'IA est parfois classée selon la faculté humaine qu'elle imite : apprentissage, raisonnement, planification, perception, etc. La classification par capacité proposée par (Schmid *et al.*, 2021) étend et détaille celle par faculté humaine : Sentir, Traiter et comprendre, Agir, Communiquer sont les 4 catégories du premier niveau. Au deuxième niveau, on a 10 capacités, par exemple Natural Language Processing (NLP) et Interaction homme-machine pour la partie Communiquer.

Une deuxième façon de classifier l'IA consiste à prendre en considération les fonctionnalités qu'elle prend en charge, par exemple en quatre catégories : l'IA réactive (sans capacité d'apprentissage), l'IA à mémoire limitée, l'IA à théorie de l'esprit (intégrant les pensées et les émotions), l'IA consciente d'elle-même (dotée de sensibilité) (Hintze, 2016).

Il existe d'autres façons de classifier, par exemple par domaine d'application, par la valeur apportée, par dimension (organisationnelle, humaine, technologique ou système d'information (SI) (Lee *et al.*, 2023), avec des guidelines pour mettre en place l'IA). Ils structurent de cette façon les antécédents, les conséquences, les défis et les guidelines pour implémenter l'IA dans les organisations. Enfin, on peut trouver des classifications selon les techniques mises en œuvre. Si l'on s'en tient à des articles très cités, mentionnons (Borges *et al*., 2021) qui structurent les domaines d'application de l'IA par ce qu'elle apporte (création de valeur) : l'aide à la décision, l'engagement du client ou de l'employé, l'automatisation, les nouveaux produits ou services. Collins *et al*. (2021) répertorient la recherche en Intelligence Artificielle pour en recenser la valeur business et les contributions dans le domaine des SI. Ils proposent une classification de l'IA en : apprentissage automatique (« machine learning »), visionique (computer vision), traitement de la langue naturelle, robotique, systèmes experts, etc. Ce qui différencie ces classifications, c'est leur granularité et/ou le périmètre plus ou moins vaste qu'elles couvrent. Certaines classifications ne couvrent en fait que l'apprentissage automatique, laissant de côté les techniques classiques de raisonnement, tels les systèmes experts. C'est ce type de classification qui nous intéresse dans cet article dans la mesure où les risques sont induits par les techniques mobilisées.

2.2. Sur les typologies de risques liés à l'IA

De nombreuses listes de risques sont aussi présentes dans la littérature. Elles ne sont pas forcément structurées ni exhaustives. Citons toutefois l'article de (McLean *et al*., 2023) qui cible l'IA générale (AGI pour Artificial General Intelligence). Il distingue les catégories de risques suivantes : perte de contrôle par l'humain, développement d'objectifs dangereux, développement d'AGI dangereux, éthique médiocre, risques existentiels. Cette liste est très spécifique, ciblant l'IA générale dont tous les experts disent qu'elle n'est pas encore opérationnelle. Certains articles étudient les risques éthiques qui sont classés selon le principe qu'ils violent : transparence, respect de la vie privée, imputabilité, équité sont les plus mentionnés (Khan *et al*., 2022). Certaines typologies ciblent un domaine, par exemple la santé : Muley *et al.* distinguent les risques liés aux données cliniques, les risques techniques et les risques socio-éthiques (Muley *et al.*, 2023). Dans (Akoka et Comyn-Wattiau, 2022), nous avons proposé une typologie des risques liés aux données, construite avec des praticiens. Elle structure les risques en trois catégories : stratégiques et réputation, légaux et réglementaires, opérationnels. Le groupe de travail AIRS (Artificial Intelligence Risk and Security)[2] propose une classification en quatre catégories : risques relatifs aux données, les attaques contre les systèmes de « machine learning », les risques liés au manque de transparence (incluant les biais), le risque de non-conformité.

Il existe des publications plus mathématiques sur l'évaluation du risque. Citons par exemple (Giudici *et al*., 2024) qui propose quatre indicateurs de risques SAFE

[2] https://www.airsgroup.ai/artificial-intelligence-governance

(pour Sustainability, Accuracy, Fairness, Explainability) et des modèles mathématiques pour les estimer. A notre connaissance, il n'existe pas d'approche structurée pour guider les décideurs dans l'identification du risque associé à l'usage des techniques d'intelligence artificielle. Mentionnons l'approche de Buehler et al. qui proposent de confronter les six catégories de risques (respect de la vie privée, sécurité, honnêteté, transparence et explicabilité, sûreté et performance, risques tiers) et les six contextes business (données, sélection et entraînement de modèle, déploiement et infrastructure, contrats et assurance, légal et réglementaire, organisation et culture) (Buehler *et al.*, 2021). Il manque cependant une opérationnalisation de la matrice à un grain plus fin. Les contextes business ne sont pas directement associés à des processus métiers. Cette publication émane d'un cabinet de conseil qui n'a pas rendu disponible plus de matériel. C'est pour combler cette lacune que nous proposons une approche décrite dans la section suivante de cet article.

3. L'approche

Dans cette section, nous décrivons notre approche d'aide à l'identification des risques induits par l'usage de l'IA pour les organisations. Nous nous appuyons sur les sciences de conception (design science) en commençant par décrire les exigences qui ont guidé nos choix dans l'élaboration de la méthode. La deuxième partie décrit le modèle conceptuel sous-jacent qui est le support des différentes étapes qui sont décrites dans la suite.

3.1. Les exigences

Notre question de recherche conduit à la proposition d'une méthodologie. Keller et Binz ont étudié les exigences auxquelles les méthodologies de conception doivent répondre (Keller et Binz, 2009). Nous avons repris et adapté celles qui nous semblaient les plus pertinentes face aux demandes des responsables qui souhaitent une aide pour identifier les risques qu'ils prennent quand ils adoptent l'IA. Ainsi, on retrouve l'exigence *d'utilité* (« usefulness »), de *compréhensibilité*, d'*apprentissage* (learnability). La *spécificité du problème* réside dans le périmètre d'application de la méthode qui doit apporter le guidage pour identifier tous les risques potentiels sans néanmoins fournir l'expertise d'évaluation chiffrée du risque. L'approche doit faciliter la *structuration* de la démarche au moyen d'étapes clairement définies et logiquement enchaînées. Elle doit aussi être *flexible* pour permettre l'évolution des contenus. Constatant l'absence d'aide structurée pour l'identification des risques induits par l'usage de l'intelligence artificielle, nous ciblons une *approche générique* qui englobe tous les types de risques, associés à chaque technique d'IA et prenant en compte tous les processus métiers. Une telle approche doit être *robuste* pour s'appliquer à différents domaines et prendre en charge de façon dynamique les nouvelles menaces au gré de l'évolution des techniques d'IA. Elle doit être *efficace* et *rentable*. Son utilisation répétée doit en faciliter l'enrichissement progressif et permettre une fertilisation croisée entre les domaines auxquels on l'applique.

L'approche doit être *semi-automatique*. En effet, elle ne peut être automatique du fait de sa complexité. Elle ne peut être manuelle du fait de son caractère fastidieux. En revanche, elle doit offrir des guides en ligne qui facilitent l'identification du risque. Elle doit aussi permettre *l'apprentissage* des concepts et des règles qui les régissent. Sa mise en œuvre doit être *facile* du fait qu'elle s'adresse en premier lieu à des décideurs non informaticiens qui sont essentiellement des managers. C'est sur la base de ces exigences que nous avons structuré l'approche décrite ci-après.

3.2. Vue d'ensemble de l'approche

Dans cette partie, nous décrivons successivement le modèle conceptuel sur lequel se fonde l'approche puis ses étapes principales.

3.2.1. Le modèle conceptuel

Ce modèle rassemble les concepts utilisés dans l'approche. Au niveau le plus élémentaire, on associe un **risque** à une **menace**, par exemple *prioriser le profit aux dépens du bien-être* est un risque associé à une menace de biais. Les risques sont regroupés en **sous-catégories** hiérarchiquement rattachées à des **catégories**. Par exemple, le risque de *prioriser le profit aux dépens du bien-être* relève de la sous-catégorie des *risques éthiques*, elle-même rattachée à la catégorie des *risques de stratégie et réputation*. Un même risque peut être rattaché à plusieurs sous-catégories. Ainsi, le risque de *discriminer via les données* relève de la sous-catégorie *Environnement-Social-Gouvernance (ESG)* de la catégorie *risques de stratégie et réputation*, mais aussi de la sous-catégorie *conformité à la loi* de la catégorie *lois et réglementations*. La menace de biais résulte de l'utilisation de la technique de traitement de la langue naturelle **(technique d'IA)**, qui peut être utilisée pour traiter automatiquement les sinistres en assurance (**processus métier**).

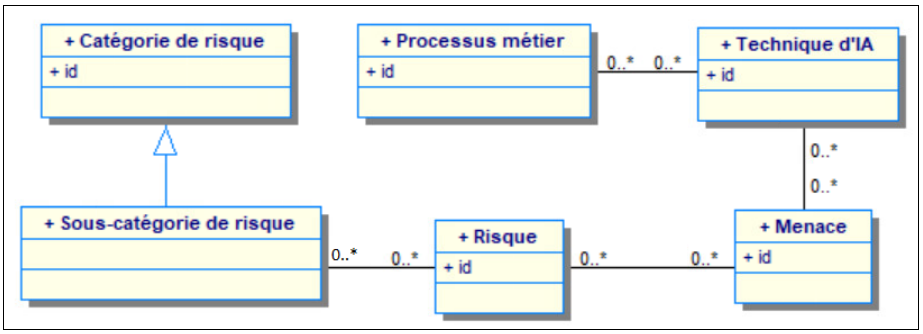

Figure 1. Le modèle conceptuel sous-jacent

3.2.2. Les étapes de l'approche

L'identification des risques liés à l'IA requiert un effort conséquent dans la mesure où il faut considérer, un par un, tous les types de risques et les confronter à tous les usages potentiels de l'IA dans le processus métier étudié. C'est la raison

pour laquelle nous proposons de procéder à une identification systématique qui soit le plus possible généralisable à différents domaines. De plus, cet effort d'identification nécessite différentes compétences (connaissance des processus métiers, compréhension des techniques d'IA et maîtrise des types de risques). En décomposant les étapes de cette identification des risques, on peut atteindre ces deux objectifs de généralisabilité et de répartition par domaine de compétences.

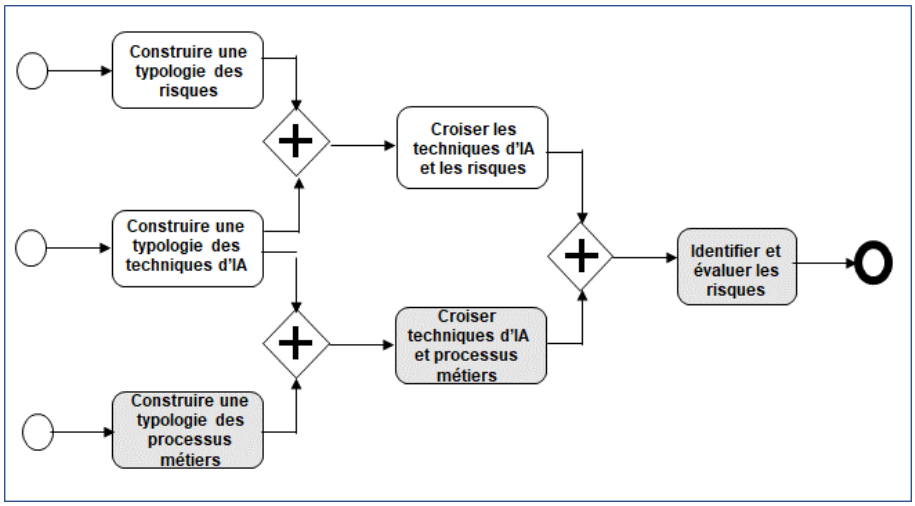

Figure 2. Les étapes de l'approche

Notre démarche comporte six étapes (Figure 2). Les tâches sur fond gris sont spécifiques aux processus métiers d'un domaine. Au premier niveau à gauche et en parallèle, on construit trois typologies respectivement des risques, des techniques d'IA et des processus métiers relatifs à un domaine. Ces typologies sont, dans une certaine mesure, indépendantes et peuvent être construites séparément. A noter qu'elles ne peuvent pas être figées parce qu'elles modélisent des domaines très évolutifs. Au deuxième niveau, suivent deux étapes de mise en correspondance (mapping). L'une de ces étapes consiste à croiser les techniques d'IA et les risques potentiels. La seconde met en relation les processus métiers et les techniques d'IA qu'ils peuvent mobiliser. La première est indépendante du secteur d'activité. En la constituant ainsi, on peut ensuite l'appliquer à différents secteurs d'activité. La seconde requiert une connaissance d'un métier ou d'un secteur d'activité. Enfin, au troisième et dernier niveau, l'effort consiste à composer les deux matrices résultats des deux étapes de mise en correspondance. La composition s'effectue sur la dimension commune et permet d'obtenir, *in fine*, une matrice croisant les risques et les processus métiers. L'intérêt de l'approche réside dans la réutilisation de l'effort d'un domaine à un autre : deux typologies génériques et une typologie métier, une matrice de croisement générique et une autre spécifique à un métier.

3.3. L'approche générique

On peut trouver de nombreuses typologies pour chacun des concepts principaux de notre approche. Toutefois, une telle typologie n'est pas pertinente dans tous les

contextes. On cherche ici à définir une approche générale de l'identification des risques liés à l'usage de l'IA, qui soit adaptée ou adaptable à n'importe quel processus métier. Les typologies sont le moyen de définir, au grain adéquat, les concepts à mettre en correspondance.

3.3.1 Typologie des risques

Rappelons qu'il n'existe pas de typologie des risques liés à l'IA qui soit reconnue comme standard. Nous avons repris et adapté celle élaborée dans (Akoka et Comyn-Wattiau, 2022) pour évaluer l'impact négatif des données et l'avons enrichie par consultation de la littérature académique et professionnelle. Parmi les trois catégories de risques (stratégie et réputation, légal et réglementaire, opérationnel), la table 1 liste ceux de la première catégorie. Chaque catégorie est ensuite décomposée en sous-catégories. La catégorie *Stratégie et réputation* comprend ainsi les sous-catégories : ESG, Ethique, Confiance et Prise de décision. L'aspect *Légal et réglementaire* est lui-même décomposé en la Conformité aux lois, la Conformité aux réglementations et à la Propriété intellectuelle. Enfin, la catégorie *Risques opérationnels* se décompose selon les trois dimensions Personnes, Processus et Technologies. Nous avons compilé la littérature pour confronter cette hiérarchie aux différents risques liés à l'usage de l'IA et avons pu, dans chaque cas, trouver une correspondance simple ou multiple entre la menace et une ou plusieurs sous-catégories de risques. Dans la table 1, la première colonne décrit la sous-catégorie de risque. La seconde colonne liste les risques relatifs à cette sous-catégorie, expliqués partiellement dans la colonne 3. Pour des raisons d'espace, la table 1 ne recense que les éléments de la catégorie risques stratégiques et de réputation. A titre d'exemple, les données personnelles sont, dans la plupart des pays, soumises à des réglementations ou des lois qui prévoient des sanctions importantes en cas de non-respect. Ainsi la conformité aux réglementations est illustrée par les risques « enfreindre le Règlement Général sur la Protection des Données (RGPD) » et « enfreindre le règlement japonais Protection of Personal Information Act (APPI) ». La typologie initiale était limitée aux risques liés aux données, alors que l'IA englobe tous les risques liés aux données mais génère des risques aussi au niveau des modèles et des systèmes (Schneider *et al*., 2023). Dans la table 1, les « nouveaux » risques sont en italiques. Par exemple, pour les risques ESG, on a identifié au moins cinq risques au lieu de trois, incluant les tensions sociales liées à toutes les craintes pour le futur de l'emploi et l'attention accrue des médias sur le sujet de l'IA. A noter que la construction de la typologie des risques a mobilisé les entités Risque, Sous-catégorie de risque et Catégorie de risque ainsi que leurs relations respectives (Figure 1). Par exemple, les sous-catégories ESG, Ethique, Confiance et Prise de décision appartiennent de manière exclusive à la catégorie Stratégie et réputation. Cela est traduit par la relation d'héritage. Le risque *Disséminer une information erronée* appartient à au moins deux sous-catégories différentes, respectivement ESG et Confiance, ce qui est traduit par la relation « appartient à » de type n:n. Certains risques appartiennent même à des catégories différentes.

Table 1. Risques stratégiques et de réputation

Sous-catégorie	Risque	Explication
ESG (Environnement-Social-Gouvernance)	Augmenter l'empreinte carbone	Les systèmes d'IA sont coûteux en temps machine
	Discriminer via les données	L'IA peut prendre des décisions biaisées ou inexactes et traiter de manière inéquitable certains groupes de personnes
	Disséminer une information erronée (sur la gouvernance)	L'IA utilisée par le site web peut générer des informations erronées
	Attirer l'attention des médias	*L'attention portée par les médias à l'IA peut amplifier les risques de réputation*
	Générer des tensions sociales	*L'automatisation grâce à l'IA fait craindre des pertes d'emploi*
Ethique	Disséminer une information fallacieuse	*L'IA peut prendre des décisions biaisées ou inexactes et traiter de manière inéquitable certains groupes de personnes*
	Fournir une information non objective sur les produits	Les systèmes d'IA utilisés pour générer des avis sur les produits peuvent être piratés
	Privilégier le profit au détriment du bien-être des clients	*Les systèmes d'IA peuvent être influencés par les préjugés personnels des personnes qui les ont créés*
Confiance	Disséminer une information erronée	*Dépendance excessive à l'égard de l'IA : un expert peut, face à de nouvelles situations se limiter aux seuls cas auxquels l'IA a également accès*
	Subir une perte d'information sensible	Les expériences négatives liées à l'IA (atteintes à la vie privée) peuvent nuire à la réputation
	Prendre des décisions singulières ou hétérogènes	Les expériences négatives liées à l'IA (décisions biaisées) peuvent nuire à la réputation
	Fournir des informations non étayées	*Les modèles d'IA sont difficiles à interpréter et compliquent l'explication des décisions aux clients et autres interlocuteurs*
Prise de décision	Prendre une décision sur la base de données erronées	Prendre des décisions erronées, entraînant des pertes financières et/ou la désaffection des clients
	Prendre une décision sur la base de données obsolètes	
	Prendre une décision convenue	*Certaines techniques d'IA ne font pas preuve de créativité (exemple : les systèmes experts)*

3.3.2. Typologie des techniques d'IA

Cet article fait suite à une présentation faite devant un public de professionnels et a permis d'aboutir à cette typologie. Ainsi, face à toutes les typologies des techniques d'IA disponibles, nous avons retenu une typologie très simple à un niveau qui nous semble, à ce stade, de nature à permettre une confrontation, d'une part, aux processus métiers susceptibles d'y recourir et, d'autre part, aux risques que ces techniques peuvent engendrer. Cette typologie comprend les six catégories suivantes :

– l'apprentissage automatique, incluant l'apprentissage profond (*deep learning*), l'apprentissage supervisé ou non, l'apprentissage par renforcement (*reinforcement learning*), etc.

– le traitement de la langue naturelle et la fouille de texte (*text mining*),

– la visionique (*computer vision*),

– la robotique ou automatisation,

– les systèmes experts (incluant les systèmes à base de règles, la représentation des connaissances, le raisonnement dans l'incertain, etc.),

– l'IA générative.

Cette typologie n'est pas une partition : certaines autres techniques sont moins répandues comme les algorithmes génétiques et n'y figurent pas, d'autres techniques se retrouvent dans plusieurs catégories. Ainsi certaines techniques de NLP font partie de l'IA générative et/ou font appel aux techniques d'apprentissage. Cette typologie mobilise l'entité Technique d'IA du modèle conceptuel (Figure 1). A noter que notre choix de typologie couvre plus que l'apprentissage automatique sans toutefois être exhaustif, notamment pour certaines techniques plus rares comme les algorithmes génétiques.

Une fois les deux typologies construites, l'approche consiste à les croiser pour faciliter l'identification des risques.

3.3.3. Mise en correspondance des techniques d'IA et des risques

L'objectif de cette étape est de remplir la matrice dont les colonnes sont les techniques d'IA (paragraphe 3.3.1) et les lignes sont les risques associés (paragraphe 3.3.2) (Table 2). Cette étape a été réalisée en analysant les informations disponibles sur les sites et la littérature scientifique traitant du sujet. L'avantage de l'approche est de pouvoir la bâtir indépendamment du domaine d'application.

Le processus de mise en correspondance vise à identifier le type de menace encourue lors de l'usage de la technique d'IA en s'aidant de la liste de tous les risques potentiels. Pour matérialiser ce risque, on caractérise son existence par la menace sous-jacente. Par exemple, le risque de discrimination via les données représente une menace qualifiée de biais lors du recours au machine learning, ou au traitement de la langue naturelle ou à la visionique. Un autre exemple est celui d'un robot qui utilise des informations erronées ou obsolètes (risque « prendre des décisions sur la base de données erronées ou obsolètes ») et peut, par là-même générer une menace de sécurité physique s'il est en interaction avec des personnes

(véhicule autonome par exemple). Cette étape matérialise l'entité Menace reliée d'une part à la technique d'IA et d'autre part à l'entité Risque (Figure 1).

Table 2. Correspondance entre techniques d'IA et risques

Technique d'IA / Risque	Machine Learning	NLP	Visionique	IA générative	Systèmes experts	Robotique
Augmenter l'empreinte carbone	Impact Environnemental					
Discriminer via les données	Biais					
Disséminer une information erronée sur la gouvernance				Altération de la réalité		
Attirer l'attention des médias	Impact Gouvernance					
Générer des tensions sociales				Impact Social		
Risque Ethique						
Disséminer une information fallacieuse	Biais					
Fournir une information non objective sur les produits	Violation de l'intégrité			Violation de l'intégrité		
Privilégier le profit au détriment du bien-être des clients	Biais					
Risque Confiance						
Disséminer une information erronée				Altération de la réalité		
Subir une perte d'information sensible		Violation de la vie privée ; perte de confidentialité				
Prendre des décisions singulières ou hétérogènes	Biais				Mauvaise déduction	
Fournir des informations non étayées	Opacité			Opacité		
Risque prise de décision						
Prendre une décision sur la base de données erronées	Mauvaise décision			Biais, opacité, mauvaise décision	Mauvaise décision	Insécurité physique
Prendre une décision sur la base de données obsolètes						
Prendre une décision convenue					Manque de créativité	

Pour des raisons d'espace, la table 2 ne contient que les risques stratégiques et de réputation. Bien entendu, cette table n'est pas exhaustive, le nombre et la nature des menaces étant très évolutifs. Dans l'avenir, on pourrait utiliser cette matrice comme structure pour une base de connaissances qui accumulerait les événements publiés sur ces menaces.

3.4. Application de l'approche aux processus métier de l'assurance

Il nous faut en premier lieu construire une typologie des processus métiers de l'assurance.

3.4.1. Typologie des processus métiers de l'assurance

Une organisation qui souhaite identifier les risques qu'elle court si elle met en place l'IA doit, au préalable, lister les processus métiers qui sont concernés. A titre d'illustration, nous avons élaboré une typologie des processus métiers du domaine de l'assurance en nous inspirant de la littérature professionnelle[3]. L'examen de celle-ci permet de recenser les processus métiers bénéficiaires de l'IA[4]. En combinant ce recensement avec la chaîne de valeur de Porter (Robben, 2014), nous avons obtenu la liste suivante : souscription de contrat, fixation des prix et évaluation des risques, service client et fidélisation, traitement des sinistres, atténuation des risques et prévention des pertes, administration des polices d'assurance, mise en conformité réglementaire, modélisation du risque et réassurance, rétention des clients, détection et prévention de la fraude.

Certains processus sont très spécifiques à l'assurance mais d'autres sont plus standards, par exemple le service client et fidélisation ou la conformité réglementaire. Cette typologie mobilise l'entité Processus métier du modèle conceptuel (Figure 1).

3.4.2. Mise en correspondance des techniques d'IA et des processus métiers

Au cours de cette étape, nous générons la matrice résultant de la mise en correspondance des processus métier, ici l'assurance, et des techniques d'IA (Table 3). Par exemple, le processus Souscription de contrat peut bénéficier de techniques de langage naturel afin de générer le texte du contrat par composition d'articles extraits d'une base de contrats-type. Il peut bénéficier aussi du machine learning qui analyse l'acceptabilité du client fondée sur ses données personnelles (classe d'âge, historique d'accidents, état de santé, revenu, etc.). Un autre exemple propre à l'assurance est celui du traitement des sinistres qui peut bénéficier de l'apport du traitement du langage naturel pour analyser les documents envoyés par le client, de robotique pour capter les images décrivant le sinistre et de visionique pour analyser ces images. Cette phase de l'approche mobilise la relation Utilise entre les entités Processus métier et Technique d'IA.

---

[3] https://acpr.banque-france.fr/sites/default/files/medias/documents/20220114_as132_transfo_numerique_assurance.pdf

[4] https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/implementing-generative-ai-with-speed-and-safety

Table 3. Correspondance entre techniques d'IA et processus de l'assurance

IA Processus	Machine Learning	NLP	Visionique	IA générative	Systèmes experts	Robo-tique
Souscription de contrat	Analyse de données	Extraction de texte				
Fixation des prix et évaluation des risques	Modélisation des facteurs de risques ; analyse des patterns			Reconnaissance de "patterns"	Estimation du risque en fonction du profil	
Service client et fidélisation	Systèmes de recommandation	Analyse de sentiments ; chatbots et assistants virtuels		Assistants virtuels ; expérience personnalisée		
Traitement des sinistres		Analyse de documents	Analyse d'image			Capture d'image
Atténuation des risques et prévention des pertes	Méthodes ensemble ; analyse prédictive					
Administration des polices d'assurance						Automatisation robotisée des processus
Conformité réglementaire		Vérification de document			Vérification via règles	
Modélisation du risque et réassurance	Analytique avancée					
Rétention des clients	Prédiction du "churn"	Communication personnalisée		Expérience personnalisée		
Détection et prévention de la fraude	Détection d'anomalies ; catégorisation des sinistres				Identification de schémas de fraude	

3.4.3. *Identification des risques*

Le but de cette phase est de fournir au décideur une matrice résultant de la composition des deux précédentes grâce aux techniques d'IA qui sont la dimension commune (Table 4). Elle comprend en ligne tous les risques triés par catégorie et en colonne les processus métiers (ici l'assurance). La table 4 n'est qu'un extrait de trois processus croisés avec six risques. Le contenu de chaque cellule a été obtenu par réinterprétation de la composition des deux matrices précédentes. Par exemple, en partant du risque « Générer des tensions sociales », la matrice de la table 2 met en

avant l'impact social de la robotique, de l'IA générative et des systèmes experts. La matrice de la table 3 repère ces techniques dans le processus de traitement des sinistres (capture d'image par robot) et dans le processus d'administration des polices d'assurance lui aussi impacté par l'automatisation par robots logiciels. La matrice de la table 4 résulte d'une réinterprétation du résultat obtenu par composition.

Table 4. Identification des risques pour les processus d'assurance

Processus **Risque**	**Souscription de contrat**	**Traitement des sinistres**	**Administration des polices d'assurance**
Risque ESG			
Discriminer via les données	Biais dans l'analyse de données ; biais dans la composition automatique de texte	Analyse biaisée des documents et des images	
Générer des tensions sociales		Tensions sociales liées au remplacement des experts d'assurance par la robotique	Tensions sociales liées au remplacement des gestionnaires de contrat par des robots
Risque éthique			
Disséminer une information fallacieuse	Biais dans l'analyse de données ; biais dans l'extraction de texte	Biais dans l'analyse des documents et des images du sinistre	
Privilégier le profit aux dépens du bien-être	Biais dans l'analyse de données		
Risque confiance			
Subir une perte d'information sensible		Divulgation d'informations privées dans l'analyse des documents et des images du sinistre	
Prendre des décisions singulières ou hétérogènes	Biais dans l'analyse des données	Mauvaise conclusion tirée de l'analyse des documents et des images	
Risque prise de décision			
Prendre une décision sur la base de données obsolètes	Non applicable		
Prendre une décision convenue			

Cette matrice peut servir à différents décideurs. Le responsable du processus métier peut utiliser le résultat pour anticiper les risques et les menaces pouvant résulter de l'utilisation de l'IA dans son champ de responsabilité. Entre différentes techniques IA, on pourrait enrichir les matrices en hiérarchisant chacun des risques. Quant au gestionnaire des risques, il se voit faciliter l'audit et la veille en repérant les techniques qui sont potentiellement génératrices de nouveaux risques et les entités concernées. Enfin, le data scientist, en charge de certaines techniques d'IA, est garant de la complétude des matrices des Tables 2 et 3 et peut assister les responsables de processus dans leurs choix de techniques, compte-tenu des risques encourus. Cette matrice peut être utilisée à différentes fins, notamment pour structurer la jurisprudence faisant état des dysfonctionnements rendus publics sur ces risques.

4. Conclusion et recherche future

Face à l'explosion de l'offre de solutions logicielles à base d'IA, les entreprises et organisations sont conscientes des opportunités mais aussi des risques. Pour ce second point, elles ne disposent pas facilement d'aide à leur identification. Nous avons présenté, dans cet article, une approche qui capitalise sur des typologies de risques, de techniques IA et de processus métiers pour faciliter l'identification de tous les risques stratégiques, juridiques et opérationnels encourus du fait de l'utilisation des techniques d'IA. L'approche fournit un moyen systématique d'analyser les conséquences de l'introduction d'une technique d'IA dans un processus. L'originalité principale est de composer une matrice générique IA*risques et une matrice spécifique IA*processus métiers. La première, une fois construite et validée, peut être enrichie par des experts du domaine de l'IA, accompagnés par des spécialistes du risque. La seconde doit être mise en place dès qu'on considère un nouveau domaine d'activité. Elle requiert les experts métiers pour cartographier les processus et peut s'appuyer sur la presse professionnelle qui met en avant les opportunités de l'IA, même sans en mentionner les risques, lesquels sont ensuite obtenus au moyen de l'autre matrice. A l'image de l'effort que les assureurs déploient pour couvrir le risque cyber, l'approche peut aussi être utilisée par eux pour proposer une offre de service à leurs clients en vue de couvrir l'usage de l'IA.

Dans la recherche future, nous prévoyons de compléter la batterie de matrices avec les outils d'évaluation mathématique des risques ainsi que les moyens d'atténuation. Une autre extension consistera à assortir les risques d'une échelle permettant de les prioriser. Pour mener plus avant la validation, nous prévoyons aussi de tester l'approche dans d'autres domaines, par exemple la logistique. Les matrices présentées dans l'article sont encore en construction et requièrent des experts de différents profils pour les valider et les faire évoluer. Enfin, le développement d'un outil semi-automatique d'aide à la décision pour faciliter le parcours des matrices et leur composition est à l'étude.

Remerciements. Les auteurs remercient les partenaires de la Chaire ESSEC Stratégie et gouvernance de l'information où cette recherche a été menée.

Bibliographie

Akoka J., Comyn-Wattiau I. (2022). Evaluation de la valeur des données - Modèle et méthode. *40ème congrès INFORSID*, Association Inforsid, Dijon, France. p.163-178.

Borges A., Laurindo J., Spinola M., Gonçalves R. Mattos C. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions, *International Journal of Information Management*, vol. 57, 102225.

Buehler K., Dooley R., Grennan L., & Singla,A. (2021). Getting to know—and manage—your biggest AI risks. *Mckinsey and Company*.

Collins C., Dennehy D., Conboy K., Mikalef P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda, *International Journal of Information Management*, vol. 60, 102383, ISSN 0268-4012.

Giudici P., Centurelli M., Turchetta S. (2024). Artificial Intelligence risk measurement, *Expert Systems with Applications*, vol. 235, 121220.

Hintze, A. (2016) Understanding the Four Types of AI, from Reactive Robots to Self-Aware Beings. *The Conversation*.

Keller A., Binz H. (2009). Requirements on engineering design methodologies. In *DS 58-2: Proceedings of ICED 09, the 17th International Conference on Engineering Design, Vol. 2, Design Theory and Research Methodology, Palo Alto, CA, USA.*

Khan A., Badshah S., Liang P., Waseem M., Khan B., Ahmad A., … & Akbar M. A. (2022). Ethics of AI: A systematic literature review of principles and challenges. In *Proceedings of the 26th Intl. Conf. on Evaluation and Assessment in Software Engineering*, p. 383-392.

Lee M., Scheepers H., Lui A., Ngai E. (2023). The implementation of artificial intelligence in organizations: A systematic literature review, *Information & Management*, vol. 60, n°5,103816, ISSN 0378-7206.

McLean S., Read G., Thompson J., Baber C., Stanton N., Salmon P. (2023). The risks associated with Artificial General Intelligence: A systematic review, *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 35, n°5, p. 649-663.

Muley A., Muzumdar P., Kurian G., & Basyal G. (2023). Risk of AI in healthcare: A comprehensive literature review and study framework. *arXiv preprint arXiv:2309.14530*.

Pereira V., Hadjielias E., Christofi M., Vrontis D. (2023). A systematic literature review on the impact of artificial intelligence on workplace outcomes: A multi-process perspective, *Human Resource Management Review*, vol. 33, n°1, 100857, ISSN 1053-4822.

Robben X. (2014). *La chaîne de valeur de Porter : Identifier la création de valeur*. 50 Minutes.

Schmid T., Hildesheim W., Holoyad, T. *et al.* (2021). The AI Methods, Capabilities and Criticality Grid. *Künstl Intell*, vol. 35, p. 425–440.

Schneider J., Abraham R., Meske C., Vom Brocke J. (2023) Artificial Intelligence Governance For Businesses, *Information Systems Management*, vol. 40 n°3, p. 229-249.