# Sécurité, intelligence artificielle et confiance : impacts et solutions pour une meilleure implémentation au sein des organisations

Fabrice Lollia

**Colloque international « intelligence artificielle et équité sociale »**
Pessac le 15,16 et 17 mai 2023
Docteur Fabrice Lollia, laboratoire Dicen idf
Fabricelollia@gmail.com

Acte de colloque

# Sécurité, intelligence artificielle et confiance : impacts et solutions pour une meilleure implémentation au sein des organisations.

# Security, Artificial Intelligence and Trust: Implications and solutions for better implementation in organisations.

# Seguridad, inteligencia artificial y confianza: repercusiones y soluciones para una mejor implantación en las organizaciones

Lollia Fabrice, laboratoire DICEN Île de France

Résumé :

La société fait face à une pervasivité sécuritaire, conduisant à une pensée solutionniste prédominante au sein des organisations. Cependant, cette confiance dans une technologie augmentée par l'IA pour résoudre les problèmes sans intervention humaine révèle des failles et pose des problèmes de légitimité, de perception, d'acceptabilité et de reconnaissance dans le domaine de la sécurité organisationnelle.

Mots clefs : Intelligence Artificielle ; Pensée Solutionniste ; Sécurité ; Technologie Augmentée ; Légitimité ; Confiance.

Summary:

Society is facing a security pervasiveness, leading to a predominant solutionist thinking within organisations. However, this reliance on AI-enhanced technology to solve problems without human intervention reveals flaws and raises issues of legitimacy, perception, acceptability and recognition in the field of organisational security.

Keywords: Artificial Intelligence; Solution Thinking; Security; Augmented Technology; Legitimacy; Trust.

Resumen:

La sociedad se enfrenta a una omnipresencia de la seguridad que conduce a un pensamiento solucionista predominante en las organizaciones. Sin embargo, esta confianza en la tecnología aumentada por IA para resolver problemas sin intervención humana revela fallos y plantea cuestiones de legitimidad, percepción, aceptabilidad y reconocimiento en el ámbito de la seguridad organizativa.

Palabras clave: Inteligencia Artificial; Pensamiento Solucionista; Seguridad; Tecnología Aumentada; Legitimidad; Confianza.

Aujourd'hui, les nouvelles technologies ont envahi le champ de la sécurité. La protection des personnes et des biens laisse désormais place à une technologie « pervasive » (Claverie *et al*., 2009).

Cette tendance à la perfection sécuritaire, due à une pensée solutionniste (Vigouroux-Zugasti, 2018), entraîne aujourd'hui l'application d'une « sécurité augmentée ». Le terme « augmenté » signifie, en l'espèce, que ce sont les performances des technologies de sécurité qui sont augmentées, alimentant une logique du « toujours plus », du « toujours mieux », mais avec toujours moins d'humains.

Nombreux sont les exemples de ce principe de sécurité augmentée, caractéristique d'une montée progressive de cette pervasivité sécuritaire. Notamment, l'insuffisance du nombre des effectifs sur le terrain est aujourd'hui palliée par l'ajout de technologies de sécurité (vidéoprotection, caméras-piétons, reconnaissance faciale, technologie antifraude) au sein des villes comme des organisations (Diard, 2018 ; Lollia, 2019 ; Diard & Dufour, 2022). Désormais, dans cette logique de perfectibilité, Les outils de sécurité sont renforcés par l'intelligence artificielle (IA).

De nombreux cas comme celui du Stade de France illustre les failles de ce déterminisme technologique basé sur la perception d'une technologie qui, parce qu'elle est augmentée par l'IA, pourrait résoudre les problèmes sans intervention humaine. Il en ressort que l'utilisation de l'IA dans le champ de la sécurité pose un problème de légitimité, de perception, d'acceptabilité (confiance, travail collaboratif) et de reconnaissance (fierté à la suite du travail accompli).

Cela nous amène aux interrogations suivantes : l'IA et la sécurité sont-elles compatibles avec la confiance de l'usager dans le domaine sécuritaire ? L'IA donne-t-elle une légitimité aux prises de décisions ? Nous tenterons d'y apporter des réponses en posant l'hypothèse que dans le champ sécuritaire, l'IA, aussi nécessaire soit-elle en apparence, soulève résolument un problème de confiance dû à des questions d'acceptabilité, d'intelligibilité et d'éthique.

Pour ce faire, nous adopterons ici un positionnement constructiviste et interprétatif.

Cette communication aura pour objet de démontrer le caractère nécessaire, sinon inévitable de l'implémentation de l'IA au sein des nouvelles technologies sécuritaires. Nous analyserons également les risques inhérents à l'IA et à la sécurité.

La création de cette culture de l'excellence amène aussi à s'interroger sur la dépendance aux résultats de ces technologies « augmentées ».

Face à toutes ces interrogations, nous proposerons non seulement des éléments de réponse, mais également des solutions éthiques concernant les biais directs (comme les discriminations) et indirects (par exemple, le risque de crédit pour une catégorie sociale non prise en compte, déformant le résultat de l'analyse finale) ou la construction des algorithmes, avec la question de leur neutralité pour corriger un biais par un autre biais.

Pour terminer, nous présenterons quelques solutions pour renforcer le trinôme IA/sécurité/confiance, notamment en nous référant au principe de l'*ethic by design.*

## I. Sécurité, intelligence artificielle et confiance : quel constat ?

### Méthodologie, concept

Nous utiliserons un positionnement constructiviste et interprétatif à partir d'une perspective d'intelligence économique mobilisant l'interdisciplinarité des sciences de l'information et de la communication, d'une part, et des informations issues d'une veille sécuritaire quotidienne alimentant nos recherches dans ce domaine depuis quelques années, d'autre part. Ce positionnement sera également complété par notre perspective expérientielle empirique de professionnel et de chercheur de terrain en recherche-action, ayant exercé pendant 22 ans au sein de la police nationale.

### Définition de l'IA et contextualisation

Dans un contexte généraliste, il existe de nombreuses définitions de l'IA, car cette notion revêt encore un aspect ambigu et flou. En Angleterre, par exemple, « *42 % des personnes interrogées ont proposé une définition plausible de l'IA, tandis que 25 % pensent qu'il s'agit de robots* » (Cave *et al*., 2019). Cela prouve que l'IA est une notion mal connue.

La littérature scientifique, quant à elle, en donne également plusieurs définitions. Par exemple, s Gamkrelidze *et al*. (2020) & Hassani *et al.* (2020),

C'est à partir de l'ensemble de ces définitions que nous formulons la nôtre : l'IA est globalement une forme d'intelligence informatique nourrie par l'augmentation technologique dans le but d'améliorer, grâce à l'imitation des facultés cognitives humaines, la prise de décision et l'apprentissage (Lollia, 2021).

Aujourd'hui, l'IA possède de multiples fonctionnalités : recherche et analyse d'informations, technologies de reconnaissance faciale, recommandations, prédictions, exécution des décisions… Elle s'applique dès lors dans de nombreux domaines. Ce large développement soulève des interrogations, et plus spécifiquement dans le champ de la sécurité, concernant par exemple la notion de confiance. En effet, dans la plupart des nouvelles technologies de l'information et de la communication (NTIC), de grandes quantités de données, notamment personnelles, sont manipulées, ce qui pose des questions légitimes en termes de sécurité, de transparence, voire de contrôle (Picard, 2017).

**Une technologie pervasive et solutionniste…**

Dans le champ sécuritaire, l'augmentation des technologies de sécurité et leur pervasivité amènent à s'interroger sur la confiance accordée à ces outils, mais aussi sur leur aspect sécuritaire et sociétal de façon générale.

Par exemple, les calculs algorithmiques de l'IA sont efficaces, mais l'absence d'intelligibilité de cette « boîte noire » entraîne certains risques, tels que celui lié à l'injection de données erronées, pouvant ainsi conduire à de mauvaises décisions sur la base de processus pourtant valides. Concernant la notion de « boîte noire » et son intelligibilité, celle-ci exige une transparence des algorithmes et renvoie à la problématique des biais volontaires ou involontaires, directs ou indirects. Ainsi, pour la prise de décision d'un algorithme, il devrait y avoir un processus de lecture transparent dans un souci de compréhension. Cela permettrait de savoir comment cette décision a été prise et donc de voir si des éléments importants sont manquants ou, au contraire, si des éléments non pertinents ont été pris en compte. On pourrait également vérifier si le traitement des informations a été biaisé dès le début par des données erronées, voire faussées dès la conception de l'algorithme sans même que l'utilisateur le sache.

Ces questions renforcent la nécessité d'un déterminisme social mais aussi d'une réflexion sur l'humain et sa complexité. D'un point de vue sécuritaire, sur le terrain, cela se traduit par des exemples concrets. On peut citer notamment les problèmes rencontrés aux États-Unis avec la reconnaissance faciale (*Le Monde*, 2020), où la question de la catégorisation ethnoculturelle a été soulevée comme un risque avéré. S'ajoute également le traitement des données aboutissant à des discriminations liées au genre, au sexe, à des raisons sociales. Il y a donc des risques dus à un usage criminel ou accidentel se basant sur le *big data,* et pouvant transformer l'algorithme en un outil discriminant, silencieux et systémique (Halevi & Moed Dr, 2012).

Par ailleurs, il est utile de souligner, sur un plan culturel cette fois-ci, que ces problématiques posent des difficultés sociales liées à la perception des gens vis-à-vis des nouvelles technologies de sécurité. En Afrique, par exemple, certains assimilent le développement de l'IA à une sorte de « cybercolonisation » (Badaoui & Najah, 2021). En France, les questions relatives à l'IA et à la reconnaissance faciale sont davantage axées sur le respect des libertés individuelles, avec une attention particulière de la part des pouvoirs publics pour projeter sur la société l'idée d'un outil de protection plutôt que de surveillance. Un autre exemple, Au Canada, de nombreuses sociétés privées travaillent sur la possibilité d'une IA prédictive dont le but est de prédire le comportement des terroristes incarcérés afin de pouvoir évaluer les chances de récidive.

Tous ces exemples montrent le caractère culturel qui détermine la perception de l'IA et son usage, le rôle essentiel de l'humain dans la gestion de ces outils de sécurité, avec la nécessité d'une approche non solutionniste, mais également le risque de dérive si « l'éthique de l'intelligibilité » n'est pas prise en compte dès le départ.

**…Qui prend le contrôle sur la prise de décision**

L'adoption d'une vision « capacitante » (Zacklad, 2018) permet de comprendre la nécessité du contrôle de l'humain sur la machine. Celle-ci s'oppose d'ailleurs à une approche transhumaniste développée par certains chercheurs, qui y voient une forme de complémentarité substitutive où la machine viendrait remplacer l'humain là où celui-ci est défaillant, comme un nouvel organe remplace celui qui ne fonctionne plus (Zacklad, 2018).

Aussi, il est utile de s'interroger sur l'impact de la prise de décision sur les aspects tactiques (à court terme) de la stratégie sécuritaire. On pense, par exemple, aux situations d'urgence sur le terrain, telle celle où un policier n'a qu'une fraction de seconde de réflexion avant d'agir. L'IA permet aujourd'hui de mettre en place des logiciels d'entraînement pour aider les policiers à réagir de façon optimale dans des situations de crise. Devons-nous, en nous appuyant sur l'IA, améliorer et automatiser nos réflexes par un entraînement constant, ou faire plutôt confiance à notre propre expérience ? Dans ce cas, à qui devons-nous imputer la faute si notre réaction n'est pas conforme aux attentes de la situation ? La confiance que nous accordons à l'IA et à son association avec l'homme ne risque-t-elle pas de nous rendre dépendants dans nos prises de décisions ? Un début de réponse peut être trouvé dans l'exemple du GPS, qui nous indique le chemin et nous est aujourd'hui indispensable pour suivre un itinéraire.

Un autre élément à prendre en compte porte sur les interactions entre l'humain et l'IA dans le champ sécuritaire. Les recherches de Reeves et Nass (1996) montrent que les attentes sociales sont les mêmes lorsque l'humain communique avec une entité artificielle. On assigne à celle-ci des règles d'interactions sociales, une situation qui peut même affecter la prise de décision. En effet, on en arrive à un point où l'homme peut déléguer certaines décisions à l'IA, même si son propre choix est meilleur en raison justement de son caractère humain (Agrawal *et al*., 2019). L'impact de l'IA sur la prise de décision soulève donc de nombreux questionnements.

Il est vrai cependant qu'en matière sécuritaire, l'IA est particulièrement **performante**. Ses utilisations sont multiples : prévoir le comportement des détenus, détecter des signaux faibles, identifier des personnes recherchées, notamment par la reconnaissance faciale, gérer des incidents grâce à une catégorisation plus précise et hiérarchisée. Prenons l'exemple des jeux Olympiques de Paris, qui vont attirer des milliers de personnes de nationalités diverses. Pour une gestion optimale des infractions et des dépôts de plaintes pour agression ou vol, notamment, les forces de l'ordre devront prendre en compte des dizaines de langues différentes. L'utilisation de l'IA et de la traduction multilingue pourra donc s'avérer nécessaire pour qu'elles puissent mener à bien leur mission.

Il en va de même pour la gestion des foules. Grâce à la vision par ordinateur, il est possible d'analyser leurs mouvements dans un cadre de régularisation et une perspective sécuritaire. Par exemple, concernant les récents incidents au Stade de France, la gestion des foules aurait pu être améliorée, notamment en termes d'évacuation des participants. L'exercice est cependant difficile : d'un côté, le cadre légal ne permet pas d'identifier les individus, mais de l'autre, il faut mettre en évidence leurs actes.

Le problème est que la méfiance alimentée par un sentiment de surveillance a jeté le doute sur les capacités de protection au sein de la société française. La peur de tomber dans les mêmes dérives que des régimes autoritaires tels que la Chine explique cette situation.

Enfin, l'IA est soumise à un risque sécuritaire dans le sens où elle peut être utilisée pour des activités sensibles telles que le pilotage de drones ou dans les systèmes de transport. Ainsi, une perte de contrôle de l'IA par l'homme, dans le cadre d'une approche exclusivement solutionniste, pourrait entraîner de graves conséquences sécuritaires.

## II. Pour quelles solutions ?

### L'éthique comme outil de cadrage ?

Dans le domaine sécuritaire, l'éthique est une dimension importante. En effet, l'introduction de l'IA a un impact sur les organisations et le comportement des acteurs (Carayol & Laborde, 2019). Elle modifie les interactions humaines et le fonctionnement des structures, en altérant progressivement, et à dose homéopathique, les avantages attendus de son utilisation (Dramba, 2019).

Aussi, à moyen terme, le risque serait que les décideurs s'appuient trop sur l'IA pour la gestion de leurs activités, ce qui pourrait fragiliser les rapports humains qui tendent à s'équilibrer lors de la transformation digitale des organisations.

Avec la pervasivité des NTIC (Claverie *et al.*, 2009) dans le champ sécuritaire et leur augmentation par l'intelligence artificielle, il est désormais indispensable de mettre en place des préconisations concernant leur conception, leur usage et leur contrôle. Ce type d'action est d'autant plus complexe que l'introduction de la dimension éthique ne doit pas freiner les avancées technologiques et les avantages sociétaux permis par l'IA. Dans cette lignée de pensée, de nombreuses approches éthiques sont développées en tenant compte de l'évolution des bonnes pratiques humaines dans le domaine de l'IA et des technologies associées, mais également en intégrant des méthodes organisationnelles afin de garantir que leur usage n'enfreint pas les principes moraux. L'*ethic by design* est alors l'approche utilisée (Béranger, 2015).

### L'exemple de l'*ethic by design* et des niveaux d'analyse pour une *ethic by evolution*

L'*ethic by design* est une approche de recherche et développement qui prend en compte les enjeux humains dans la conception des applications d'IA. Son objectif est double : la rendre compréhensible pour le plus grand nombre, et mieux encadrer le code des programmeurs afin d'interdire les algorithmes liberticides et indignes (Bensoussan, 2020). Cette problématique est prise au sérieux au niveau international, comme en Allemagne,

où le respect des exigences éthiques en matière d'IA est important pour la plupart des citoyens (Kieslich *et al*., 2022).

Ensuite, en écho aux recherches de Dignum (2018), il est possible, par le biais d'une analyse conceptuelle à trois niveaux (éthique dès la conception, éthique dans la conception, éthique de la conception), de trouver son efficacité par une conception contrôlée tout au long de la chaîne de fabrication.

Enfin, l'éthique par évolution apparaît comme une approche intéressante au regard de la protection de l'IA et de l'humain dans le champ de la sécurité. Son objectif est de créer, déployer, utiliser et surveiller des dispositifs algorithmiques innovants, à travers une démarche systémique qui intervient à tous les niveaux de l'organisation (techniciens, ingénieurs, dirigeants, personnes concernées par l'usage de l'IA). Ainsi favorise-t-elle le respect ou la mise en place de règles éthiques grâce à une vue d'ensemble dans le cycle de vie de l'IA, permettant alors une mise en conformité permanente avec les textes de référence et notamment les textes juridiques.

**La collaboration : un système gagnant-gagnant**

La collaboration est également un système permettant d'adopter l'IA dans une perspective sécuritaire. Le partage des connaissances entre le secteur public et le secteur privé reste essentiel à la création de normes et de règlements nécessaires à la bonne implémentation de l'IA dans la société et à la sécurité de tous. C'est ce qui permettra d'asseoir la confiance des utilisateurs et de conserver un aspect sécuritaire global afin de protéger la société.

**Conclusion**

L'utilisation de l'IA connaît une progression fulgurante dans la vie de tous les jours et dans de nombreux secteurs, notamment pour améliorer la sécurité publique à l'occasion par exemple de grands événements. Toutefois, la confiance envers cet outil technologique reste un sujet majeur. Les risques de manipulation de l'IA peuvent entraîner des résultats erronés, avec des conséquences discriminantes importantes dues à un manque de transparence et d'intelligibilité. Il est donc devenu indispensable de veiller à une conception éthique de l'IA, laquelle est rendue possible grâce à des concepts comme l'*ethic by evolution*. Il faut alors bien comprendre qu'aujourd'hui, même si l'IA apporte des améliorations sécuritaires importantes, elle doit être associée à des principes éthiques forts mais également être envisagée selon un continuum entre développeur, gouvernement et entreprise.

**Fabrice Lollia, laboratoire DICEN Île-de-France**

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Exploring the impact of artificial Intelligence : Prediction versus judgment. *Information Economics and Policy*, *47*, 1-6. https://doi.org/10.1016/j.infoecopol.2019.05.001

Badaoui, S., & Najah, R. (2021). *Intelligence Artificielle et Cyber-colonisation : Implications sur l'Afrique/Artificial Intelligence and Cyber-colonization : Implications for Africa*.

Bensoussan, A. (2020). Quelle régulation juridique pour l'intelligence artificielle ? *Enjeux numériques. Intelligences humaines et artificielles. Quelles interactions?*, *12*. http://annales.org/enjeux-numeriques/2020/en-12-12-20.pdf

Béranger, J. (2015). *Les systèmes d'information en santé et l'éthique : D'Hippocrate à e-ppocr@ te*. ISTE Group.

Carayol, V., & Laborde, A. (2019). Les organisations malades du numérique. *Communication et organisation*, *56*, 11-17. https://doi.org/10.4000/communicationorganisation.8207

Cave, S., Coughlan, K., & Dihal, K. (2019). « Scary Robots » : Examining Public Responses to AI. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 331-337. https://doi.org/10.1145/3306618.3314232

Claverie, B., Lespinet-Najib, V., & Fouillat, P. (2009). Pervasion, transparence et cognition augmentée. *Revue des Interactions Humaines Médiatisées (RIHM)= Journal of Human Mediated Interactions*, *10*(2), 85-99.

Diard, C. (2018). Psychological acceptance of organizational video surveillance. *Human Systems Management*, *37*(1), 105-115. https://doi.org/10.3233/HSM-17113

Diard, C., & Dufour, N. (2022). Technologies de contrôle : Un enjeu organisationnel de lutte contre la fraude interne ? *Management & Avenir*, *130*, 65-89.

Dignum, V. (2018). Ethics in artificial intelligence : Introduction to the special issue. *Ethics and Information Technology*, *20*(1), 1-3. https://doi.org/10.1007/s10676-018-9450-z

Dramba, M. (2019). Mise en place d'un dispositif numérique de gestion du temps : De l'agencement managérial à la trahison. *Communication et organisation*, *56*, 33-46. https://doi.org/10.4000/communicationorganisation.8256

Gamkrelidze, T., Zouinar, M., & Barcellini, F. (2020). Gamkrelidze, T., Zouinar, M. & Barcellini, F. (Sous presse). Les anciens enjeux des « nouveaux » systèmes d'Intelligence Artificielle dans les activités professionnelles. In *Les transformations digitales à l'épreuve de l'activité et des salariés : Comprendre et accompagner les mutations technologiques.*

Halevi, G., & Moed Dr, H. F. (2012). The evolution of big data as a research and scientific topic : Overview of the literature. *Research trends*, *1*(30), 2.

Hassani, H., Silva, E. S., Unger, S., TajMazinani, M., & Mac Feely, S. (2020). Artificial Intelligence (AI) or Intelligence Augmentation (IA) : What Is the Future? *AI*, *1*(2), 143-155. https://doi.org/10.3390/ai1020008

Kieslich, K., Keller, B., & Starke, C. (2022). Artificial intelligence ethics by design. Evaluating public perception on the importance of ethical design principles of artificial intelligence. *Big Data & Society*, *9*(1), 205395172210929. https://doi.org/10.1177/20539517221092956

Le Monde. (2020). *Etats-Unis : Un Américain noir arrêté à tort à cause de la technologie de reconnaissance faciale*. https://www.lemonde.fr/international/article/2020/06/24/un-americain-noir-arrete-a-tort-a-cause-de-la-technologie-de-reconnaissance-faciale_6044073_3210.html

Lollia, F. (2019). Organisation en milieu hostile : L'effet de la géolocalisation sur l'organisation en milieu terroriste. *Journal of Human Mediatised Interactions/Revue des Interactions Humaines Médiatisées*, *20*(2). https://doi.org/10.5281/ZENODO.4587093

Lollia, F. (2021). *Digital transformation : A literature review of the integration of artificial intelligence into the company's organisational strategy.* An International and Interdisciplinary Perspective on Digital Transformation: The Case of Developing and Emerging Economies./Workshop international" Une perspective internationale et interdisciplinaire sur la transformation numérique".

Picard, F. (2017). Une éthique du numérique centrée sur les citoyens et orientée vers des solutions techniques. *Livre Blanc ADEL, Vade-mecum sur la gouvernance des traitements de données numériques*.

Reeves, B., & Nass, C. I. (1996). *The media equation : How people treat computers, television, and new media like real people and places*. CSLI Publications ; Cambridge University Press.

Vigouroux-Zugasti, E. (2018). Morozov Evgeny, 2014. Pour tout résoudre, cliquez ici : L'aberration du solutionnisme technologique: Limoges : Fyp éditions. ISBN 978-2-36405-115-7. 22,50 €. *Revue française des sciences de l'information et de la communication*, *13*. https://doi.org/10.4000/rfsic.3573

Zacklad, M. (2018). *Intelligence Artificielle : Représentations et impacts sociétaux*. 15 pages. https://halshs.archives-ouvertes.fr/halshs-02937255