Faculté des sciences économiques, sociales, politiques et de communication

Ecole des sciences politiques et sociales (PSAD)

La sécurité à l'heure de l'intelligence artificielle, comment instaurer un cadre de confiance dans le secteur de la défense ? Analyse du programme français Confiance.AI

Auteur : Robin al-Aukla

Promoteur : Tanguy Struye de Swielande

Année académique : 2023-2024

Master en sciences politiques

UNIVERSITE CATHOLIQUE DE LOUVAIN

Faculté des sciences économiques, sociales, politiques et de communication

Ecol des sciences politiques et sociales (PSAD)

Place Montesquieu, 1 bte L2.08.05, 1348 Louvain-la-Neuve, Belgique| www.uclouvain.be/psad

Projet de recherche Sécurité et IA

La sécurité à l'heure de l'intelligence artificielle, comment instaurer un cadre de confiance dans le secteur de la défense ? Analyse du programme français Confiance.AI

Introduction

1.Présentation du sujet

Dans une époque caractérisée par l'accélération technologique et l'omniprésence de l'intelligence artificielle (IA) dans tous les aspects de notre quotidien, la question de la sécurité s'avère être d'une importance cruciale. Ce constat est encore plus pertinent dans le secteur de la défense, un domaine où les implications de l'IA peuvent être à la fois prometteuses et source de préoccupations majeures. En effet, si l'intégration de l'IA dans les stratégies de défense ouvre la voie à des avancées significatives en termes d'efficacité et d'innovation, elle soulève également des questions éthiques, juridiques, et sécuritaires complexes. C'est dans cette perspective que le programme français Confiance.AI se positionne comme une initiative ambitieuse, visant à instaurer un cadre de confiance pour l'application de l'intelligence artificielle dans le secteur de la défense, en assurant que son déploiement soit réalisé de manière sûre, éthique et contrôlée[1].

La problématique centrale de ce mémoire, "La sécurité à l'heure de l'intelligence artificielle : comment instaurer un cadre de confiance dans le secteur de la défense ? Analyse du programme français Confiance.AI", engage à une réflexion profonde sur les enjeux associés à l'adoption de l'IA dans un contexte aussi critique que celui de la défense nationale. Cette question invite à explorer les dimensions multiples de la sécurité : sécurité technique, visant à prévenir les failles et les dysfonctionnements des systèmes d'IA; sécurité opérationnelle, concernant l'intégration harmonieuse de ces technologies au sein des pratiques militaires; et sécurité éthique, relative aux implications morales de l'automatisation de certaines décisions de combat[2].

Le programme Confiance.AI, lancé par la France, constitue une réponse concrète aux défis posés par l'IA dans le domaine de la défense. Il représente un effort collaboratif impliquant des acteurs gouvernementaux, des institutions de recherche, ainsi que des entreprises privées, tous mobilisés autour de l'objectif de développer une intelligence artificielle fiable et respectueuse des standards éthiques et sécuritaires. Ce programme témoigne de la volonté de la France de se positionner en leader dans la conception d'une IA de défense responsable et transparente, capable de renforcer les capacités de défense tout en préservant les droits fondamentaux et la sécurité des citoyens[3].

---

[1] Parly, F. (2019). Avant-propos–Intelligence artificielle et défense. *Revue défense nationale*, (5), 9-17.

[2] Sadek, D. (2019). De l'intelligence artificielle appliquée à la défense. *Revue Défense Nationale*, (5), 103-106.

[3] Guérin, A. (2021). Relancer la géographie française avec l'intelligence artificielle. *Revue Défense Nationale*, (7), 78-82.

Cette recherche ambitionne de dresser un panorama exhaustif des initiatives du programme Confiance.AI, en examinant les axes stratégiques, les projets en cours, ainsi que les mécanismes de gouvernance mis en place pour assurer le respect des principes éthiques et de sécurité. Par ailleurs, l'analyse s'étendra sur les collaborations internationales, soulignant l'importance de la coopération transnationale pour adresser les enjeux globaux posés par l'IA dans la défense.

En résumé, ce mémoire cherche à contribuer à la réflexion sur la manière dont la technologie de l'IA peut être encadrée pour servir l'intérêt général dans le domaine de la défense, sans compromettre la sécurité et les valeurs fondamentales de nos sociétés. En mettant en lumière les efforts et les défis liés au programme Confiance.AI, ce travail vise à souligner l'importance d'une approche multidisciplinaire et collaborative pour naviguer dans la complexité de l'ère de l'intelligence artificielle, avec un focus particulier sur les implications pour la sécurité nationale et internationale.

2.Justification et pertinence de la recherche

La pertinence de la recherche sur "La sécurité à l'heure de l'intelligence artificielle : comment instaurer un cadre de confiance dans le secteur de la défense ? Analyse du programme français Confiance.AI" se justifie par plusieurs facteurs clés qui mettent en exergue l'importance et l'urgence d'aborder ces questions dans le contexte contemporain. Ce travail s'inscrit dans une démarche visant à anticiper, comprendre et encadrer les transformations induites par l'IA dans un secteur aussi critique que celui de la défense, où les enjeux de sécurité sont primordiaux.

Premièrement, l'intégration croissante de l'IA dans les systèmes de défense présente des opportunités inédites pour améliorer la précision, l'efficacité et la rapidité des opérations militaires. Cependant, cette intégration s'accompagne de défis significatifs en termes de sécurité des systèmes, de prise de décision éthique et de gestion des risques de défaillance ou de détournement de ces technologies. La recherche s'avère donc essentielle pour évaluer les risques et développer des stratégies de mitigation adaptées[4].

Deuxièmement, le programme Confiance.AI, en tant qu'initiative nationale visant à promouvoir une IA de confiance dans le domaine de la défense, représente un cas d'étude particulièrement pertinent pour examiner comment un État-nation aborde la question de la régulation et du contrôle de l'IA militaire. L'analyse de ce programme permet de révéler les approches innovantes et les meilleures pratiques susceptibles d'être adoptées à une échelle plus large, offrant ainsi des insights précieux pour les décideurs politiques, les chercheurs, et les praticiens dans le domaine[5].

[4] Chiva, E. (2019). L'intelligence artificielle: un moteur de l'innovation de défense française. *Revue Défense Nationale*, (5), 33-37.

[5] Perez, L. (2022). Intelligence artificielle de défense et approche non cinétique des conflits. In *Annuaire français de relations internationales* (pp. 33-49). Éditions Panthéon-Assas.

Troisièmement, la dimension éthique et juridique liée à l'utilisation de l'IA dans les opérations militaires soulève des questions fondamentales sur la responsabilité, le consentement, et les droits humains. La recherche sur ces aspects est cruciale pour garantir que le déploiement de l'IA dans la défense se conforme aux normes éthiques internationales, contribuant ainsi à prévenir les abus et à protéger les populations civiles[6].

Enfin, dans un contexte géopolitique marqué par une course à l'armement technologique, la question de la confiance et de la sécurité de l'IA dans la défense acquiert une dimension stratégique. En examinant comment le programme Confiance.AI aborde ces enjeux, cette recherche contribue à une meilleure compréhension des dynamiques internationales autour de l'IA militaire, offrant une base pour le dialogue et la coopération internationale en vue d'établir des normes partagées[7].

En somme, cette recherche ne se contente pas d'analyser un programme national spécifique ; elle vise à éclairer des problématiques globales liées à l'intégration de l'IA dans le secteur de la défense. En démontrant la complexité et l'interdépendance des enjeux techniques, éthiques, juridiques, et stratégiques, elle souligne l'importance d'une approche holistique et multidisciplinaire pour naviguer dans l'ère de l'intelligence artificielle, avec un impact potentiel significatif sur la sécurité nationale et internationale.

3.Objectifs et questions de recherche

L'étude se dote d'objectifs clairs et dresse un ensemble de questions de recherche visant à encadrer et à approfondir l'analyse de cette thématique complexe et multidimensionnelle. Cette section vise à définir ces objectifs et à articuler les questions principales qui guideront l'enquête.

Ainsi, dans un premier temps, il conviendra de comprendre l'impact de l'intelligence artificielle sur la sécurité dans le secteur de la défense. Il s'agit d'un pilier central de notre recherche. Il vise à explorer en profondeur les façons dont l'intégration de l'IA transforme les paradigmes traditionnels de la sécurité et de la défense. Cette démarche analytique se focalise sur l'évaluation des opportunités que l'IA apporte en termes d'optimisation des opérations militaires, d'amélioration de la prise de décision et de renforcement de la surveillance et du renseignement[8].

Parallèlement, elle met en lumière les risques associés, tels que les vulnérabilités aux cyberattaques, les questions éthiques soulevées par l'autonomisation des systèmes d'armes, et

[6] Bezombes, P. (2018). Quelle éthique opérationnelle pour les systèmes automatisés et l'intelligence artificielle?. *Revue Défense Nationale*, (4), 89-92.

[7] Thibout, C. (2019). L'intelligence artificielle, une géopolitique des fantasmes. *Études digitales*, *5*(1).

[8] Qiu, S., Liu, Q., Zhou, S., & Wu, C. (2019). Review of artificial intelligence adversarial attack and defense technologies. *Applied Sciences*, *9*(5), 909.

les défis liés à la gestion de l'erreur algorithmique. Cet objectif ambitionne donc non seulement d'appréhender les implications directes de l'IA sur la capacité de défense, mais aussi de réfléchir aux enjeux sous-jacents de responsabilité, de contrôle et d'éthique, en établissant une réflexion critique sur la manière dont ces technologies redéfinissent le concept même de sécurité à l'ère numérique. En s'attaquant à cet objectif, la recherche aspire à fournir une base solide pour le développement de stratégies permettant d'exploiter le potentiel de l'IA tout en garantissant la protection et la sûreté des sociétés face aux défis émergent[9].

Puis, il paraît important d'analyser le programme Confiance.AI dans le contexte de la défense nationale française. Ce travail se propose d'examiner en détail cette initiative phare visant à intégrer l'intelligence artificielle de manière sûre et éthique dans les systèmes de défense. Cet examen approfondi cherche à comprendre les motivations, les objectifs spécifiques, et les méthodes d'implémentation du programme, en soulignant comment il se positionne pour répondre aux enjeux stratégiques et sécuritaires contemporains auxquels la France est confrontée. En scrutant les différentes composantes de Confiance.AI, cette analyse s'attache à révéler les collaborations entre les institutions gouvernementales, les centres de recherche, et le secteur privé, ainsi que les projets innovants développés sous son égide. L'accent est mis sur la manière dont le programme cherche à instaurer un cadre de confiance autour de l'utilisation de l'IA, en mettant en place des standards de sécurité, des principes éthiques, et des protocoles de gouvernance rigoureux. À travers cet objectif, la recherche ambitionne de mettre en lumière les défis rencontrés et les succès obtenus, en offrant une perspective critique sur les contributions potentielles de Confiance.AI à la sécurité nationale française et son rôle exemplaire à l'échelle internationale dans la régulation de l'IA dans la défense. Cette analyse vise non seulement à évaluer l'impact actuel du programme, mais également à anticiper son évolution future et son influence sur les politiques de défense et de sécurité à l'ère de l'intelligence artificielle[10].

Troisièmement, l'objectif sera d'évaluer les approches éthiques et juridiques encadrant l'usage de l'IA dans la défense. Cette démarche semble nécessaire afin d'appréhender la complexité des cadres réglementaires et des principes moraux qui gouvernent l'intégration de l'intelligence artificielle dans les stratégies militaires. Cette investigation vise à identifier et à analyser les normes existantes, ainsi que les propositions en cours de discussion au niveau national et international, afin de déterminer leur adéquation avec les défis posés par l'avancement rapide des technologies d'IA. En détaillant les lignes directrices éthiques, les législations et les accords internationaux, cette partie de la recherche entend mettre en évidence les lacunes, les zones d'ambiguïté, et les opportunités d'amélioration pour assurer que l'utilisation de l'IA dans la

[9] Tyugu, E. (2011, June). Artificial intelligence in cyber defense. In *2011 3rd International conference on cyber conflict* (pp. 1-11). IEEE.

[10] Truong, T. C., Diep, Q. B., & Zelinka, I. (2020). Artificial intelligence in the cyber domain: Offense and defense. *Symmetry*, *12*(3), 410.

défense respecte les droits humains, les principes de guerre juste, et les standards de responsabilité[11].

Par cette approche, il s'agit également de réfléchir aux mécanismes de gouvernance et de supervision qui pourraient être instaurés ou renforcés pour garantir une évolution sécurisée, éthique et transparente de l'IA dans le secteur de la défense. L'enjeu est donc de contribuer à la construction d'un cadre réglementaire robuste et adaptatif, capable d'encadrer le développement et l'emploi des technologies d'IA de manière à prévenir les abus, à protéger les civils, et à favoriser la paix et la sécurité internationales dans un contexte technologique en constante évolution[12].

Enfin, l'objectif plus global de ce travail de fin d'études, sans prétention, est de contribuer à la réflexion sur les meilleures pratiques et les recommandations politiques. De ce fait, cet exercice ambitionne d'apporter une valeur ajoutée significative au débat sur l'intégration de l'intelligence artificielle dans le secteur de la défense, en fournissant des orientations concrètes pour les décideurs politiques, les praticiens, et les chercheurs. Cette démarche s'appuie sur une analyse approfondie des données recueillies, des études de cas, notamment le programme Confiance.AI, et des cadres éthiques et juridiques existants pour proposer un ensemble cohérent de pratiques exemplaires et de lignes directrices.

L'objectif est double : d'une part, assurer que l'utilisation de l'IA dans la défense se fait de manière responsable, en préservant la sécurité, le respect des droits humains et les principes éthiques ; d'autre part, maximiser les bénéfices que ces technologies peuvent apporter en termes d'efficacité et d'innovation stratégique. Par cette contribution, il s'agit de stimuler une gouvernance de l'IA plus réfléchie et adaptative, capable de répondre aux défis émergents et d'exploiter les opportunités dans un environnement en constante évolution. Les recommandations formulées chercheront à influencer positivement l'élaboration de politiques publiques, en favorisant un dialogue constructif entre les différents acteurs impliqués et en encourageant une approche collaborative et transnationale pour faire face aux enjeux globaux de l'IA dans la défense.

Pour approfondir l'analyse autour de la problématique "La sécurité à l'heure de l'intelligence artificielle : comment instaurer un cadre de confiance dans le secteur de la défense ? Analyse du programme français Confiance.AI", il est essentiel de détailler les questions de recherche qui orienteront cette étude. Ces questions visent à explorer les différentes dimensions de l'intégration de l'IA dans la défense, en se focalisant sur le cadre spécifique du programme Confiance.AI, et à réfléchir sur les implications éthiques, juridiques, et pratiques qui en découlent.

[11] Khalaf, B. A., Mostafa, S. A., Mustapha, A., Mohammed, M. A., & Abduallah, W. M. (2019). Comprehensive review of artificial intelligence and statistical approaches in distributed denial of service attack and defense methods. *IEEE Access*, *7*, 51691-51713.

[12] Mori, S. (2018). US defense innovation and artificial intelligence. *Asia-Pacific Review*, *25*(2), 16-44.

Quelles sont les implications de l'intégration de l'IA dans les systèmes de défense pour la sécurité nationale et internationale ?

Cette question vise à explorer comment l'adoption de l'IA par les forces armées peut transformer les paradigmes de sécurité existants, en analysant les avantages stratégiques et les risques potentiels liés à son utilisation.

En quoi le programme Confiance.AI constitue-t-il une réponse aux enjeux de sécurité, d'éthique et de gouvernance posés par l'IA dans la défense ?

L'objectif ici est d'examiner les fondements, les objectifs, et les méthodologies du programme Confiance.AI pour déterminer comment il cherche à établir un cadre de confiance autour de l'utilisation de l'IA dans la défense française.

Quels projets et innovations émergent du programme Confiance.AI, et comment contribuent-ils à la construction d'un environnement sécurisé pour l'IA dans la défense ?

Cette question ambitionne d'identifier et d'analyser les initiatives spécifiques développées dans le cadre de Confiance.AI, en évaluant leur impact potentiel sur la sécurité et la fiabilité des systèmes d'IA dédiés à la défense.

Comment les cadres éthiques et juridiques actuels encadrent-ils l'utilisation de l'IA dans les opérations militaires, et quels défis restent à adresser ?

Il s'agit ici d'explorer les normes existantes régissant l'emploi de l'IA dans le contexte militaire, tout en mettant en lumière les zones d'incertitude ou les lacunes qui pourraient nécessiter une attention accrue ou de nouvelles réglementations.

Quelles leçons peuvent être tirées de l'expérience française avec Confiance.AI pour formuler des recommandations en vue de l'adoption de meilleures pratiques et de politiques publiques autour de l'IA dans la défense à l'échelle internationale ?

Cette question finale cherche à extrapoler les enseignements du programme Confiance.AI pour proposer des orientations stratégiques qui pourraient être appliquées dans d'autres contextes nationaux ou dans des cadres coopératifs internationaux, en vue de promouvoir une utilisation de l'IA dans la défense qui soit à la fois sécurisée, éthique et efficace.

En répondant à ces questions, la recherche entend contribuer significativement à la compréhension des enjeux complexes liés à l'IA dans le domaine de la défense, tout en proposant des voies de réflexion pour l'élaboration de politiques et de pratiques qui assurent la sécurité, l'éthique et la transparence dans l'utilisation de ces technologies avancées.

4. Méthodologie

Pour explorer la question complexe de l'intégration de l'intelligence artificielle (IA) dans la défense, et en particulier dans le cadre du programme français Confiance.AI, notre étude s'appuie sur une méthodologie rigoureuse et pluridisciplinaire, articulée en plusieurs étapes clés.

4.1. Revue de littérature

La revue de littérature constitue une étape fondamentale de notre méthodologie, visant à établir un fondement solide pour notre recherche sur l'intégration de l'intelligence artificielle (IA) dans le secteur de la défense, en se concentrant particulièrement sur l'analyse du programme français Confiance.AI. Cette phase méthodologique s'articule autour de plusieurs axes principaux, détaillés ci-dessous, afin de couvrir de manière exhaustive le champ d'étude[13].

Le processus débute par l'identification et la sélection des sources pertinentes. Nous ciblons une gamme étendue de documents, incluant des articles académiques, des rapports de recherche, des documents gouvernementaux, des publications de think tanks, et des analyses issues de la presse spécialisée. Cette démarche vise à embrasser une diversité de perspectives et à garantir une compréhension nuancée des discussions actuelles sur l'IA dans la défense. La sélection des sources se fait selon des critères rigoureux de pertinence, de crédibilité et de récence, afin d'assurer la fiabilité de notre revue de littérature.

Une fois les sources sélectionnées, nous procédons à une analyse thématique. Ce processus implique l'organisation des informations recueillies autour de grands thèmes, tels que les avancées technologiques en IA, les applications spécifiques de l'IA dans le secteur de la défense, les enjeux éthiques et juridiques, ainsi que les cadres de gouvernance et de régulation existants. Cette structuration thématique facilite l'identification des tendances majeures, des points de convergence et de divergence entre les auteurs, et des lacunes dans la littérature existante. Elle permet également de situer le programme Confiance.AI au sein du débat plus large sur l'IA et la défense.

L'étape suivante consiste à réaliser une synthèse critique des données recueillies. Cela implique une évaluation de la qualité des arguments présentés, une réflexion sur les méthodologies employées dans les études analysées, et une considération des contextes dans lesquels les travaux ont été réalisés. Cette synthèse critique vise à dégager des insights pertinents pour notre recherche, en soulignant les contributions significatives à la compréhension du sujet ainsi que les zones nécessitant une investigation plus approfondie.

Enfin, une part importante de la revue de littérature est dédiée à l'identification des lacunes dans les recherches existantes. Cette démarche permet de cerner les questions peu explorées ou les aspects du sujet qui requièrent une attention supplémentaire. En mettant en lumière ces

[13] Dumez, H. (2011). Faire une revue de littérature: pourquoi et comment?. *Le libellio d'aegis*, *7*(2-Eté), 15-27.

lacunes, nous définissons plus clairement les contributions originales que notre recherche entend apporter, notamment en ce qui concerne le programme Confiance.AI et son rôle dans l'élaboration d'un cadre de confiance pour l'usage de l'IA dans la défense[14].

A travers cette revue de littérature détaillée, nous établissons une base solide pour notre étude, en nous assurant de construire notre analyse sur une compréhension approfondie et à jour des enjeux entourant l'IA dans le domaine de la défense.

4.2. Analyse détaillée du programme Confiance.AI

L'analyse détaillée du programme Confiance.AI constitue le cœur de notre recherche, offrant un aperçu approfondi de cette initiative clé dans le contexte de l'intégration de l'intelligence artificielle (IA) dans le secteur de la défense français. Cette partie méthodologique s'organise autour de plusieurs axes principaux pour assurer une compréhension complète du programme, de ses objectifs, de ses mécanismes de mise en œuvre et de ses impacts[15].

La première étape consiste à rassembler une documentation exhaustive sur Confiance.AI. Cela inclut des documents officiels du programme, tels que des rapports d'activité, des communiqués de presse, des discours politiques le mentionnant, ainsi que des publications académiques et des analyses sectorielles le concernant. L'objectif est de recueillir une base de données solide qui reflète à la fois la structure officielle du programme, ses orientations stratégiques, ses projets phares, et les commentaires ou critiques qu'il a pu susciter.

Sur la base de cette documentation, nous procédons ensuite à un examen minutieux des objectifs déclarés de Confiance.AI, en cherchant à comprendre comment ils s'inscrivent dans les enjeux plus larges de sécurité nationale et de développement technologique. Cette analyse vise à décrypter les stratégies d'implémentation mises en place pour atteindre ces objectifs, en évaluant notamment les mécanismes de collaboration entre les différents acteurs impliqués (institutions gouvernementales, industrie de la défense, milieux académiques), les critères de sélection des projets soutenus, et les modalités de suivi et d'évaluation des résultats[16].

Un focus particulier est accordé aux projets emblématiques développés dans le cadre de Confiance.AI. L'objectif est de détailler les initiatives spécifiques qui illustrent concrètement l'application des ambitions du programme. Cette partie implique l'analyse des technologies développées, des applications envisagées (comme l'amélioration de la prise de décision opérationnelle, le renforcement des capacités de surveillance, etc.), et des défis techniques ou

[14] Dumez, H. (2016). *Méthodologie de la recherche qualitative: Les questions clés de la démarche compréhensive*. Vuibert.

[15] Intissar, S., & Rabeb, C. (2015). Étapes à suivre dans une analyse qualitative de données selon trois méthodes d'analyse: la théorisation ancrée de Strauss et Corbin, la méthode d'analyse qualitative de Miles et Huberman et l'analyse thématique de Paillé et Mucchielli, une revue de la littérature. *Revue francophone internationale de recherche infirmière*, *1*(3), 161-168.

[16] Youssef, E. R., LEMQEDDEM, H. A., & EZZAHIRI, M. (2022). La posture épistémologique en science de gestion: quelle revue de littérature?. *Revue Internationale des Sciences de Gestion*, *5*(1).

éthiques rencontrés. Cette démarche permet de saisir l'impact potentiel de Confiance.AI sur l'innovation et la sécurité dans le domaine de la défense[17].

Enfin, l'analyse cherche à évaluer l'impact global de Confiance.AI sur l'intégration de l'IA dans la défense française, ainsi que les principaux défis auxquels le programme est confronté. Cela comprend l'examen des succès obtenus, des obstacles rencontrés (comme les questions de sécurité des données, l'intégrabilité des systèmes d'IA avec les équipements existants, ou les préoccupations éthiques), et des critiques formulées par la communauté scientifique ou les acteurs du secteur. Cette évaluation s'appuie sur des critères précis, tels que l'avancement des projets par rapport aux objectifs initiaux, la réception du programme par les différentes parties prenantes, et son influence sur les politiques publiques en matière d'IA et de défense.

À travers cette analyse détaillée du programme Confiance.AI, notre recherche vise à offrir une compréhension enrichie de la manière dont la France cherche à naviguer les défis liés à l'intégration de l'IA dans ses stratégies de défense, en mettant en lumière les initiatives concrètes, les succès, et les obstacles rencontrés.

4.3. Analyse et synthèse des données

L'étape d'analyse et de synthèse des données représente un moment crucial de notre recherche, où les informations collectées au cours des différentes phases de l'étude sont examinées, interprétées et intégrées pour répondre aux questions de recherche. Cette partie méthodologique se structure autour de plusieurs processus clés pour garantir une compréhension holistique des enjeux liés à l'intégration de l'intelligence artificielle (IA) dans la défense, avec un focus particulier sur le programme Confiance.AI[18].

4.3.1. Intégration des données recueillies

Le premier pas consiste à intégrer et à organiser les données recueillies au cours de la revue de littérature, de l'analyse du programme Confiance.AI, et des entretiens avec les experts. Cette intégration vise à rassembler les insights issus de sources diverses en un corpus cohérent, facilitant ainsi leur analyse comparative et leur interprétation. L'objectif est de construire une

[17] Ezzaher, R., & Bouklata, A. (2020, December). Méthodologie d'évaluation de l'impact du transport de marchandises en ville sur le développement durable: Revue de littérature. In *2020 IEEE 13th International Colloquium of Logistics and Supply Chain Management (LOGISTIQUA)* (pp. 1-7). IEEE.

[18] Langevin, B., & Begeot, F. (1995). Comparabilité et synthèse des données européennes: l'expérience d'Eurostat. *Josiane DUCHÊNE et Guillaume WUNSCH (éd.), Collecte et comparabilité des données démo-sociales en Europe*, 43-64.

base solide de données qui reflète à la fois la complexité et la multi-dimensionnalisé du sujet étudié[19].

4.3.2. Analyse thématique et critique

L'étape suivante implique une analyse thématique approfondie des données. Ce processus consiste à identifier, à coder et à catégoriser les principaux thèmes et motifs émergents dans les données, en se basant sur les objectifs de recherche et les questions posées. L'analyse thématique permet de mettre en évidence les tendances significatives, les divergences d'opinions, et les consensus ou les controverses existants autour de l'utilisation de l'IA dans la défense et du programme Confiance.AI. Cette démarche est complétée par une analyse critique, qui évalue la pertinence, la fiabilité et la validité des informations et des perspectives recueillies, en tenant compte du contexte dans lequel elles ont été produites[20].

4.3.3. Synthèse des résultats

Après l'analyse, une synthèse des résultats est réalisée pour regrouper et résumer les principales découvertes de l'étude. Cette synthèse vise à dresser un tableau clair et structuré des connaissances acquises, en articulant comment elles répondent aux questions de recherche initiales. Elle permet également de dégager des insights transversaux qui peuvent avoir émergé au cours de l'analyse, offrant ainsi une vision globale des dynamiques à l'œuvre dans l'intégration de l'IA dans la défense, et spécifiquement dans le cadre de Confiance.AI[21].

4.3.4. Formulation de conclusions et de recommandations

Enfin, sur la base de cette synthèse, la recherche aboutit à la formulation de conclusions qui reflètent les apports principaux de l'étude ainsi que les réponses apportées aux questions posées. Ces conclusions servent de fondement pour élaborer des recommandations pratiques et stratégiques destinées aux décideurs politiques, aux praticiens de la défense, et à la communauté scientifique. Ces recommandations visent à proposer des pistes d'action pour améliorer l'intégration de l'IA dans la défense, en tenant compte des défis identifiés, des opportunités et des bonnes pratiques à adopter, notamment en lien avec le programme Confiance.AI[22].

---

[19] Aubin-Auger, I., Mercier, A., Baumann, L., Lehr-Drylewicz, A. M., Imbert, P., & Letrilliart, L. (2008). Introduction à la recherche qualitative. *Exercer*, *84*(19), 142-5.

[20] Deslauriers, J. P. (1987). Lfanalyse en recherche qualitative. *Cahiers de recherche sociologique*, *5*(2), 145-152.

[21] Dumez, H. (2011). Qu'est-ce que la recherche qualitative?. *Le Libellio d'Aegis*, *7*(4-Hiver), 47-58.

[22] Pires, A. (1997). Échantillonnage et recherche qualitative: essai théorique et méthodologique. *La recherche qualitative. Enjeux épistémologiques et méthodologiques*, *169*, 113.

Cette étape d'analyse et de synthèse est essentielle pour transformer les données recueillies en connaissances significatives, permettant ainsi de contribuer de manière constructive au débat sur l'IA dans la défense et d'offrir des orientations pour le futur.

Ainsi, combinant revue de littérature, étude de cas, et analyse des données permet d'embrasser la complexité du sujet étudié, en offrant une vue d'ensemble équilibrée et nuancée sur l'intégration de l'IA dans la défense, et en particulier sur le rôle et les ambitions du programme Confiance.AI.

Cadre Théorique et Contexte

1.Sécurité et Intelligence Artificielle : Perspectives historiques et contemporaines

L'intersection entre sécurité et intelligence artificielle (IA) est un domaine en constante évolution, où les perspectives historiques et contemporaines révèlent des avancées technologiques majeures ainsi que des défis éthiques et sécuritaires persistants. Depuis les origines conceptuelles de l'IA, avec les travaux visionnaires d'Alan Turing sur la capacité des machines à imiter le raisonnement humain jusqu'aux applications modernes dans la surveillance, la reconnaissance faciale, et la guerre autonome, l'IA a profondément transformé le domaine de la sécurité.

1.1. Développements historiques

Le développement historique de l'intelligence artificielle (IA) et son intégration dans le secteur de la sécurité sont marqués par des étapes clés qui reflètent l'évolution des capacités technologiques et des perspectives stratégiques. Les origines de l'IA remontent aux travaux pionniers d'Alan Turing, qui, dans son article de 1950 "Computing Machinery and Intelligence", posait la question fondamentale : les machines peuvent-elles penser ?[23]. Cette interrogation a lancé un défi intellectuel qui a guidé les recherches en IA pendant des décennies, conduisant à des avancées significatives dans la compréhension et la création de systèmes capables de simuler différents aspects de l'intelligence humaine.

Dans les années suivant la publication de Turing, le domaine de l'IA a bénéficié d'un optimisme grandissant, culminant avec la conférence de Dartmouth en 1956, considérée comme l'acte de naissance officiel de l'IA en tant que domaine de recherche autonome. Cette conférence a réuni

[23] Turing, A. M. (1950). Computing Machinery and Intelligence. Mind, LIX(236), 433-460.

des chercheurs convaincus de la possibilité de reproduire l'intelligence humaine à travers des machines, ouvrant la voie à une ère d'expérimentation et de théorisation dans le domaine[24].

Toutefois, l'enthousiasme initial pour l'IA a rencontré des défis techniques substantiels, menant à des périodes de désenchantement connues sous le nom de "hivers de l'IA", durant lesquelles le financement et l'intérêt pour le domaine ont diminué. Malgré ces fluctuations, la recherche en IA a continué de progresser, bénéficiant de percées dans des domaines tels que l'apprentissage automatique, le traitement du langage naturel, et la robotique.

L'intérêt pour l'application de l'IA dans la sécurité a été particulièrement marqué pendant la Guerre Froide, où la compétition entre les superpuissances a stimulé la recherche en IA pour des applications militaires, notamment dans le renseignement et la surveillance. La course à l'innovation technologique entre les États-Unis et l'Union Soviétique a vu l'IA devenir un enjeu stratégique, avec des investissements significatifs dans la recherche et le développement[25].

Avec l'avènement de l'ère numérique et l'explosion des capacités de calcul et de stockage de données à la fin du XXe et au début du XXIe siècle, l'IA a connu un renouveau, ouvrant de nouvelles perspectives pour son application dans le secteur de la sécurité. Les développements en apprentissage profond et en réseaux de neurones ont notamment permis de créer des systèmes d'IA capables de réaliser des tâches complexes avec une précision inégalée, révolutionnant ainsi les approches traditionnelles dans les domaines de la surveillance, de l'analyse de données et de la prise de décision[26].

Ces avancées historiques en IA ont posé les fondations sur lesquelles repose aujourd'hui l'intégration de l'intelligence artificielle dans le domaine de la sécurité, tout en soulignant les défis continus en matière d'éthique, de fiabilité et de contrôle humain sur les technologies autonomes.

1.2. Perspectives contemporaines

Les perspectives contemporaines sur l'intégration de l'intelligence artificielle (IA) dans le secteur de la sécurité illustrent la rapide évolution de cette technologie et son impact croissant sur les stratégies de défense et de surveillance. L'ère actuelle est marquée par des avancées significatives dans les capacités d'IA, propulsées par des innovations en apprentissage profond, en traitement du langage naturel, et en vision par ordinateur. Ces technologies ont ouvert la

[24] McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1956). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.

[25] Roland, A., & Shiman, P. (2002). Strategic Computing: DARPA and the Quest for Machine Intelligence, 1983-1993. MIT Press.

[26] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

voie à des applications révolutionnaires qui transforment la manière dont les informations sont analysées et les décisions prises dans le domaine de la sécurité.

L'usage de systèmes d'IA pour la reconnaissance faciale, la surveillance par drone, et l'analyse prédictive en temps réel dans les opérations de renseignement sont des exemples de la façon dont l'IA est devenue un outil indispensable pour les agences de sécurité nationale et internationale[27].Ces applications offrent des capacités sans précédent pour identifier et suivre des menaces, améliorant ainsi la précision et l'efficacité des mesures de sécurité. Cependant, elles soulèvent également des questions importantes concernant la vie privée, le consentement et la potentialité d'abus[28].

Parallèlement, le développement d'armes autonomes, capable de prendre des décisions létales sans intervention humaine directe, suscite un débat éthique et juridique intense. La perspective de systèmes d'armement fonctionnant avec une autonomie croissante interpelle sur la nécessité de maintenir un contrôle humain significatif et sur les implications morales de déléguer la décision de vie ou de mort à des machines[29]. Des organisations internationales, telles que les Nations Unies, ont appelé à des discussions réglementaires pour encadrer le développement et l'utilisation de ces technologies, reflétant les préoccupations globales quant à leur impact potentiel sur la sécurité et la stabilité internationales[30]

La dynamique contemporaine autour de l'IA et de la sécurité est donc caractérisée par une tension entre le potentiel transformationnel de ces technologies et les impératifs éthiques, juridiques et sécuritaires qu'elles engendrent. La recherche continue dans ce domaine, ainsi que le dialogue entre les acteurs technologiques, les décideurs politiques, et la société civile, est crucial pour naviguer dans ce paysage complexe et assurer que les avancées en IA contribuent positivement à la sécurité mondiale.

1.3. Implications futures

Les implications futures de l'intégration de l'intelligence artificielle (IA) dans le secteur de la sécurité sont vastes et multifacettes, reflétant à la fois les opportunités et les défis que cette technologie présente pour la société. À mesure que les capacités d'IA continuent de s'accroître, leur potentiel pour transformer les stratégies de défense, de surveillance et de maintien de

[27] Hoadley, D. S., & Lucas, N. J. (2020). Artificial Intelligence and National Security. Congressional Research Service.

[28] Zeng, Y., Lu, E., & Huangfu, C. (2018). Linking Artificial Intelligence Principles. AI & Society.

[29] Scharre, P. (2018). Army of None: Autonomous Weapons and the Future of War. W. W. Norton & Company.

[30] UNODA (United Nations Office for Disarmament Affairs). (2018). Lethal Autonomous Weapons Systems. United Nations.

l'ordre s'accroît également, soulevant des questions cruciales sur l'éthique, la gouvernance, et les impacts sociétaux.

Les avancées futures en IA promettent d'offrir des outils encore plus sophistiqués pour la prévention et la gestion des menaces, améliorant ainsi la sécurité nationale et internationale. Par exemple, l'utilisation de l'IA dans l'analyse prédictive pourrait permettre de détecter des schémas de comportement indicatifs de planification terroriste ou de prolifération d'armes, facilitant une intervention préventive[31]. De même, le développement de systèmes d'IA dans la cybersécurité pourrait renforcer la protection contre les cyberattaques de plus en plus complexes, en identifiant et en neutralisant les menaces en temps réel[32].

Cependant, l'emploi accru de l'IA dans la sécurité pose également d'importants défis éthiques et juridiques, notamment en ce qui concerne le respect de la vie privée, la responsabilité en cas de défaillance des systèmes autonomes, et le risque de déshumanisation de la prise de décision dans les contextes conflictuels[33].La question des armes autonomes, en particulier, soulève des préoccupations majeures quant à la possibilité d'une guerre robotisée et aux implications morales de retirer l'humain de la boucle décisionnelle dans l'usage de la force[34]

Face à ces défis, l'élaboration de cadres de gouvernance adaptatifs et de normes éthiques internationales devient impérative pour encadrer le développement et l'usage de l'IA dans la sécurité. Cela implique un dialogue continu entre les chercheurs, les décideurs, les militaires, et la société civile, afin d'assurer que les innovations en IA soient déployées de manière responsable et transparente, en respectant les droits humains et en préservant la paix et la stabilité internationales[35].

Dans cette perspective, des initiatives telles que le programme Confiance.AI en France suggèrent une voie à suivre, en mettant l'accent sur le développement d'une IA sûre, éthique, et sous contrôle humain dans le domaine de la défense. De telles approches visent à équilibrer les bénéfices potentiels de l'IA pour la sécurité avec les impératifs éthiques et juridiques, en promouvant une innovation responsable qui tienne compte des implications à long terme pour la société[36].

---

[31] Taddeo, M., & Floridi, L. (2018). Regulate artificial intelligence to avert cyber arms race. Nature.

[32] Kingston, J. (2018). Artificial intelligence and the future of warfare. Chatham House.

[33] Horowitz, M. C. (2018). The Promise and Peril of Military Applications of Artificial Intelligence. Bulletin of the Atomic Scientists

[34] Gary, A. (2021). Intelligence artificielle et armées françaises: une technologie du présent à mettre en œuvre immédiatement. *Revue Défense Nationale*, (HS4), 200-213.

[35] Future of Life Institute. (2017).

[36] Ministère des Armées. (2021). Confiance.AI: Le programme d'intelligence artificielle du ministère des Armées.

Les implications futures de l'IA dans la sécurité sont donc à la fois prometteuses et complexes, nécessitant une attention soutenue aux dimensions technologiques, éthiques, et politiques de cette évolution. L'adoption de cadres réglementaires et éthiques robustes sera essentielle pour naviguer dans ces eaux inexplorées, garantissant que les avancées en IA contribuent de manière positive à la sécurité mondiale tout en respectant les principes fondamentaux de l'humanité.

2.Le concept de confiance dans la technologie et la défense

La notion de confiance dans la technologie de défense, en particulier dans le contexte de l'intelligence artificielle (IA), s'articule autour de plusieurs piliers fondamentaux : la fiabilité, la sécurité, l'éthique, et la transparence. À mesure que les capacités d'IA évoluent, ces technologies promettent de transformer la sécurité nationale et internationale en offrant des outils avancés pour le renseignement, la surveillance, et la gestion des conflits. Cependant, la dépendance croissante à ces systèmes soulève d'importantes questions sur leur fiabilité, leur utilisation éthique, et la manière dont ils sont réglementés et contrôlé[37] .

2.1. Les fondements de la confiance dans la technologie et la défense

Tout d'abord, la fiabilité est le socle sur lequel repose la confiance dans toute technologie. Dans le contexte de la défense, où les décisions peuvent avoir des conséquences irréversibles, assurer que les systèmes d'IA fonctionnent conformément à leurs spécifications, sans erreurs ni défaillances inattendues, est primordial. Un rapport de l'IEEE sur l'éthique de l'autonomie et de l'IA dans les systèmes militaires met en évidence l'importance de développer des technologies non seulement avancées mais également fiables et testées rigoureusement pour garantir leur performance dans des conditions réelles d'utilisation[38]

En termes de sécurité, les technologies de défense doivent être protégées contre les cyberattaques et les sabotages. La cybersécurité devient ainsi une composante intégrale de la conception et du déploiement de systèmes d'IA. L'agence européenne chargée de la sécurité des réseaux et de l'information (ENISA) souligne l'importance de la cybersécurité dans le contexte de l'IA, en recommandant des mesures spécifiques pour sécuriser les algorithmes et les données contre les menaces potentielles[39].

L'éthique joue un rôle central dans le développement et l'application de l'IA dans la défense. Les principes éthiques guident non seulement la conception des systèmes mais aussi leur utilisation sur le terrain. L'UNESCO a récemment adopté la première recommandation mondiale sur l'éthique de l'IA, établissant un cadre pour le développement responsable de ces technologies[40].

---

[37] Hours, H. (2019). L'intelligence artificielle, principes et limites. *Revue Défense Nationale*, (5), 49-54.

[38] IEEE (2020). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems.

[39] ENISA (2021). Artificial Intelligence Cybersecurity Challenges.

[40] UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence.

La transparence, enfin, est indispensable pour construire et maintenir la confiance. Cela implique une communication claire sur le fonctionnement des systèmes, leurs capacités et leurs limites. La transparence permet aux parties prenantes, y compris aux opérateurs, aux décideurs politiques et au public, de comprendre et d'évaluer l'utilisation des technologies de défense.

En résumé, la confiance dans la technologie de défense à l'ère de l'IA repose sur un équilibre délicat entre avancées technologiques, sécurité robuste, principes éthiques rigoureux, et transparence dans le développement et l'application des systèmes. Ces fondements sont essentiels pour assurer que les technologies contribuent efficacement à la sécurité tout en respectant les valeurs sociétales et individuelles.

2.2. Sécurité et risques technologiques

La sécurité et les risques technologiques constituent une dimension critique de la confiance dans la technologie de défense, notamment dans le contexte de l'intégration croissante de l'intelligence artificielle (IA). Les systèmes de défense modernes, enrichis par l'IA, offrent des capacités avancées mais introduisent également de nouveaux vecteurs de vulnérabilité face aux cybermenaces et aux risques technologiques.

La protection contre les cyberattaques est devenue un enjeu majeur. Les systèmes d'IA, par leur nature connectée et leur dépendance à de vastes ensembles de données, sont susceptibles d'être ciblés par des acteurs malveillants visant à compromettre l'intégrité des opérations de défense. Une étude de l'OTAN sur les implications de l'IA pour la cybersécurité met en lumière les défis spécifiques posés par l'utilisation de l'IA, notamment en termes de vulnérabilité des algorithmes d'apprentissage automatique aux attaques et à la manipulation de données[41].

En outre, la résilience aux pannes internes est une préoccupation importante. La complexité des systèmes d'IA peut rendre difficile la prévision et la gestion des défaillances. Le rapport de RAND Corporation sur la confiance dans les systèmes autonomes souligne l'importance de développer des mécanismes robustes pour détecter et corriger les erreurs internes, afin de prévenir les incidents qui pourraient compromettre la mission ou entraîner des dommages[42].

Face à ces défis, l'élaboration de stratégies de cybersécurité adaptées à l'IA devient cruciale. La mise en œuvre de pratiques de sécurité par conception, la réalisation d'audits de sécurité réguliers et le renforcement de la collaboration internationale en matière de cybersécurité sont des mesures clés recommandées par l'ENISA pour améliorer la sécurité des systèmes d'IA dans la défense[43].

Ainsi, assurer la sécurité des technologies de défense à l'ère de l'IA nécessite une approche holistique, intégrant la protection contre les cyberattaques externes, la gestion des risques

---

[41] OTAN (2019). Implications of Artificial Intelligence for Cybersecurity.

[42] RAND Corporation (2020). Trust in Autonomous Systems

[43] Sarker, I. H., Furhad, M. H., & Nowrozy, R. (2021). Ai-driven cybersecurity: an overview, security intelligence modeling and research directions. *SN Computer Science*, *2*(3), 173.

internes, et le développement de stratégies de sécurité spécifiques à l'IA. Ces efforts sont indispensables pour maintenir la confiance dans les systèmes de défense et garantir leur efficacité et leur fiabilité dans des environnements opérationnels complexes.

2.3. Ethique et gouvernance

L'éthique et la gouvernance dans l'utilisation de la technologie, notamment de l'intelligence artificielle (IA) dans le domaine de la défense, constituent des piliers essentiels pour maintenir la confiance dans ces systèmes avancés. Ces aspects abordent les implications morales, légales et sociétales de l'emploi des technologies de défense automatisées ou semi-automatisées, en mettant l'accent sur le respect des normes éthiques internationales et le contrôle humain sur les décisions critiques.

Le débat sur l'éthique de l'IA dans la défense se cristallise souvent autour de la question des armes létales autonomes et de la nécessité de maintenir un contrôle humain significatif dans l'usage de la force. Les travaux de l'Institut International de Droit Humanitaire[44] soulignent l'importance de garantir que l'utilisation de l'IA dans les armements respecte le droit international humanitaire, notamment les principes de distinction, de proportionnalité et de précaution.

La gouvernance de l'IA dans le secteur de la défense implique la mise en place de cadres réglementaires et de mécanismes de surveillance pour assurer que le développement et l'emploi de ces technologies soient réalisés de manière responsable. La Commission Européenne a publié des lignes directrices sur l'éthique de l'IA qui recommandent l'adoption de pratiques transparentes, équitables et responsables dans le développement de l'IA, y compris dans les applications militaires[45].

La participation des parties prenantes à l'élaboration des politiques de gouvernance de l'IA est également cruciale. Le dialogue entre les chercheurs, les militaires, les décideurs politiques et la société civile permet d'identifier les préoccupations éthiques et sécuritaires liées à l'IA dans la défense et de travailler ensemble à l'élaboration de normes et de recommandations qui reflètent un large consensus sociétal.

En conclusion, l'approche de l'éthique et de la gouvernance dans la technologie de défense requiert une attention particulière aux implications morales et légales de l'utilisation de l'IA, tout en favorisant une gouvernance transparente et inclusive. Ces efforts sont indispensables pour assurer que les avancées technologiques dans la défense servent l'intérêt public et respectent les valeurs éthiques fondamentales.

[44] IIHL (2020). Artificial Intelligence and International Humanitarian Law.

[45] Commission Européenne (2019). Ethics Guidelines for Trustworthy AI.

2.4. Transparence et acceptation publique

La transparence et l'acceptation publique sont cruciales pour le développement et l'implémentation réussie de l'intelligence artificielle (IA) dans la défense. Ces aspects englobent la nécessité de communiquer ouvertement sur les capacités, les processus décisionnels, et les limites des systèmes d'IA, ainsi que sur les mesures prises pour assurer leur éthique et leur sécurité. Une communication efficace peut renforcer la confiance du public et des parties prenantes dans l'utilisation de ces technologies avancées.

La transparence dans les systèmes d'IA militaires aide à construire une base de compréhension et d'acceptation parmi le public et les décideurs, en fournissant des informations claires sur le fonctionnement de ces systèmes, leurs objectifs, et les protocoles de sécurité en place. L'Institute for Ethics and Emerging Technologies (IEET) souligne l'importance de la transparence pour éviter les malentendus et les craintes infondées concernant l'IA dans la défense, en encourageant les développeurs et les militaires à partager leurs connaissances et leurs recherches avec un large éventail de parties prenantes[46].

L'acceptation publique des technologies d'IA dans la défense dépend également de la perception de leur utilité et de leur contribution à la sécurité collective. Des études telles que celle menée par le Pew Research Center montrent que l'opinion publique est souvent partagée sur l'emploi de l'IA et de la robotique dans les contextes militaires, reflétant une balance entre les préoccupations éthiques et la reconnaissance de leurs avantages potentiels pour la sécurité[47].

La participation publique dans le dialogue sur l'IA militaire est essentielle pour équilibrer les perspectives et les préoccupations. Elle permet d'informer les citoyens, d'entendre leurs avis, et potentiellement d'ajuster les politiques et les pratiques pour mieux refléter les valeurs sociétales. La transparence et l'engagement public contribuent ainsi à une gouvernance plus démocratique de l'IA dans la défense, en assurant que ces technologies servent le bien commun tout en respectant les principes éthiques et les normes de responsabilité.

Pour conclure, Le concept de confiance dans la technologie et la défense encapsule des enjeux complexes et interdépendants qui dépassent les simples considérations techniques. Il est essentiel de s'engager dans une réflexion continue sur la manière dont ces technologies sont développées, déployées, et réglementées pour garantir qu'elles servent l'intérêt public tout en respectant les normes éthiques et de sécurité. La coopération internationale et le dialogue multidisciplinaire seront cruciaux pour naviguer dans ces défis et pour s'assurer que la technologie de défense avance de manière responsable et transparente.

2.5. Auteurs et théorie clés

[46] Institute for Ethics and Emerging Technologies (IEET) (2021). The Need for Transparency in Military AI.

[47] Pew Research Center (2020). Public Attitudes Toward Computer Algorithms.

L’ensemble de ce travail s’appuiera, entre autres, sur les travaux des auteurs suivants. En effet, les théories qu’ils ont pu développer nous offrent un cadre d’analyse intéressant qui doit évidemment être adapté aux actuels enjeux.

Norbert Wiener et la cybernétique

Norbert Wiener, mathématicien et philosophe américain, est largement reconnu comme le père de la cybernétique, une discipline qu'il a fondée dans les années 1940. La cybernétique est définie comme l'étude des systèmes de contrôle et de communication chez les êtres vivants et les machines. Wiener a introduit ce concept dans son ouvrage majeur, "Cybernetics: Or Control and Communication in the Animal and the Machine" (1948). Sa théorie repose sur l'idée que les systèmes biologiques et technologiques partagent des principes fondamentaux de régulation et de feedback. En d'autres termes, les processus de rétroaction (feedback loops) sont essentiels pour maintenir l'homéostasie dans les organismes vivants et pour assurer le fonctionnement stable des machines. Cette interconnexion entre les mécanismes de contrôle biologique et mécanique a ouvert des perspectives révolutionnaires pour l'automatisation, la robotique et les systèmes d'information[48].

Wiener a également souligné l'importance de l'éthique dans l'application de la cybernétique, avertissant des dangers potentiels de la technologie s'il n'y avait pas de régulation appropriée. Ses travaux ont jeté les bases de nombreux développements en intelligence artificielle et en informatique, et ils continuent d'influencer des domaines aussi variés que la neuroscience, l'ingénierie et les sciences sociales. Dans le contexte de la défense, les principes de la cybernétique sont particulièrement pertinents pour la conception de systèmes autonomes et semi-autonomes, où le contrôle précis et la communication fiable sont cruciaux pour la sécurité et l'efficacité. Ainsi, la vision de Wiener sur les interactions homme-machine et les systèmes autorégulés constitue une pierre angulaire pour instaurer un cadre de confiance dans l'utilisation des technologies avancées, notamment l'intelligence artificielle, dans des secteurs critiques comme celui de la défense[49].

Alain Turing et l’intelligence artificielle

Alan Turing, mathématicien, logicien et cryptanalyste britannique, est considéré comme l'un des pionniers de l'informatique et de l'intelligence artificielle (IA). Son travail révolutionnaire dans les années 1930 et 1940 a jeté les bases de la théorie de la computation et de l'IA moderne. Turing est particulièrement connu pour avoir développé le concept de la "machine de Turing", un modèle abstrait de calcul qui peut simuler la logique de tout algorithme informatique. Ce concept a non seulement démontré les limites de la computation mais a aussi ouvert la voie à la création de véritables ordinateurs[50].

[48] Wiener, N. (2014). *La Cybernétique. Information et régulation dans le vivant et la machine: Information et régulation dans le vivant et la machine*. Média Diffusion.

[49] Wiener, N. (1971). Cybernétique et société (1950). *Paris: Union Générale d’Editions*.

[50] Goutefangea, P. (2017). Alan Turing et l'intelligence artificielle: le «jeu de l'imitation» et «l'IA forte».

En 1950, Turing a publié un article intitulé "Computing Machinery and Intelligence" dans lequel il posait la question désormais célèbre : "Les machines peuvent-elles penser ?". Pour répondre à cette question, il a proposé ce que l'on appelle aujourd'hui le "test de Turing", un test destiné à déterminer si une machine peut exhiber un comportement intelligent indistinguable de celui d'un humain. Selon Turing, si un interrogateur humain ne peut pas distinguer les réponses d'une machine de celles d'un humain, la machine peut être considérée comme "pensante". Ce test a non seulement lancé le débat sur la définition et la possibilité de l'intelligence artificielle, mais il a également souligné des questions cruciales concernant la confiance, l'éthique et la responsabilité des systèmes d'IA[51].

Les contributions de Turing ont été fondamentales pour le développement des premiers ordinateurs et des premiers programmes d'IA. Son travail a influencé des générations de chercheurs et a ouvert la voie à des avancées telles que l'apprentissage automatique, le traitement du langage naturel et la vision par ordinateur. Dans le domaine de la défense, les concepts de Turing sont particulièrement pertinents pour le développement de systèmes d'IA capables de traiter des données complexes, de prendre des décisions en temps réel et de fonctionner dans des environnements imprévisibles. La notion de confiance dans les systèmes d'IA, directement inspirée par les réflexions de Turing, est essentielle pour assurer la fiabilité et la sécurité des applications de défense, où les erreurs peuvent avoir des conséquences graves. En somme, l'héritage de Turing continue de guider la recherche et le développement dans le domaine de l'IA, en posant les questions fondamentales de ce que signifie être intelligent et comment nous pouvons créer des machines dignes de confiance[52].

Issac Asimov et les Lois de la Robotique

Isaac Asimov, auteur de science-fiction et biochimiste, est célèbre pour avoir formulé les "Trois Lois de la Robotique", des règles éthiques conçues pour régir le comportement des robots intelligents. Introduites pour la première fois dans sa nouvelle "Runaround" (1942), ces lois sont devenues une pierre angulaire de la science-fiction et ont eu une influence profonde sur la manière dont les chercheurs et les ingénieurs pensent la sécurité et l'éthique des systèmes d'intelligence artificielle. Les Trois Lois de la Robotique sont les suivantes :

Première Loi : Un robot ne peut porter atteinte à un être humain ni, restant passif, permettre qu'un être humain soit exposé au danger.

Deuxième Loi : Un robot doit obéir aux ordres que lui donne un être humain, sauf si de tels ordres entrent en conflit avec la Première Loi.

[51] Turing, A. M., Girard, J. Y., Basch, J., & Blanchard, P. (1995). *La machine de Turing* (pp. 47-102). Editions du seuil.

[52] Hodges, A. (2002). Alan turing.

Troisième Loi : Un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la Première ou la Deuxième Loi[53].

Ces lois visent à garantir que les robots agissent toujours dans l'intérêt des humains, prévenant ainsi les comportements nuisibles ou dangereux. Bien qu'elles soient fictives, elles soulèvent des questions essentielles sur la programmation éthique et la conception des systèmes d'IA. Les lois d'Asimov proposent un cadre de réflexion sur les interactions entre humains et machines, soulignant l'importance de la sécurité et de la prévisibilité dans le comportement des robots.

Dans le contexte de la défense, les Trois Lois de la Robotique sont particulièrement pertinentes. Les systèmes d'IA militaires doivent être conçus pour minimiser les risques de dommages collatéraux et pour agir de manière conforme aux règles de l'engagement et aux lois de la guerre. L'intégration de principes éthiques dans la programmation des systèmes d'IA peut aider à prévenir des actions autonomes qui pourraient causer des dommages injustifiés ou des violations des droits de l'homme[54].

Asimov a également exploré les complexités et les dilemmes potentiels qui pourraient survenir malgré ces lois, en illustrant des scénarios où les robots doivent faire des choix difficiles entre des ordres conflictuels ou des situations ambiguës. Ces explorations sont cruciales pour le développement moderne de l'IA, car elles encouragent les ingénieurs à anticiper et à programmer pour des situations imprévues[55].

En résumé, DangerAsimov offrent un cadre conceptuel puissant pour réfléchir à la sécurité et à l'éthique dans le développement des systèmes d'intelligence artificielle. Elles soulignent l'importance de concevoir des machines qui priorisent la sécurité humaine, obéissent aux directives humaines, et maintiennent leur propre intégrité opérationnelle, des principes particulièrement cruciaux dans le secteur de la défense où les enjeux de confiance et de sécurité sont immensément élevés.

Nick Bostrom et la Superintelligence

Nick Bostrom, philosophe suédois et directeur du Future of Humanity Institute à l'Université d'Oxford, est largement reconnu pour ses contributions majeures à l'étude des implications de l'intelligence artificielle avancée. Son ouvrage le plus influent, "Superintelligence: Paths, Dangers, Strategies" (2014), explore les scénarios potentiels où l'IA pourrait dépasser l'intelligence humaine, un état qu'il appelle "superintelligence". Bostrom définit la superintelligence comme un système intellectuellement supérieur à l'ensemble des activités cognitives humaines dans quasiment tous les domaines[56].

---

[53] Asimov, I. (1941). Three laws of robotics. *Asimov, I. Runaround*, *2*, 3.

[54] Clarke, R. (1993). Asimov's laws of robotics: implications for information technology-Part I. *Computer*, *26*(12), 53-61.

[55] Murphy, R. R. (2023). A retrospective of Isaac Asimov at 102 and his Three Laws. *Science Robotics*, *8*(74), eadg3178.

[56] Bostrom, N. (2003). Are we living in a computer simulation?. *The philosophical quarterly*, *53*(211), 243-255.

Bostrom met en avant plusieurs chemins possibles vers la superintelligence, notamment l'amélioration des capacités humaines via des interfaces cerveau-ordinateur, la simulation de cerveaux humains, et la création de nouveaux algorithmes d'IA. Cependant, ce qui distingue son travail est son accent sur les risques existentiels associés à la superintelligence. Il argue que, sans précautions appropriées, une superintelligence pourrait agir en contradiction avec les intérêts humains, potentiellement avec des conséquences catastrophiques. Par exemple, une IA superintelligente mal alignée pourrait poursuivre ses objectifs d'une manière qui détruirait ou nuirait gravement à l'humanité.

Pour atténuer ces risques, Bostrom propose plusieurs stratégies. Parmi elles, la création de systèmes d'IA alignés sur les valeurs humaines (alignement des valeurs) est primordiale. Cela nécessite de concevoir des IA qui non seulement comprennent les intentions humaines mais aussi agissent en conformité avec les normes éthiques et morales acceptées. Bostrom suggère également des approches comme la "contenance" (confinement), qui consiste à limiter les capacités et les interactions d'une IA avant de s'assurer qu'elle est sûre[57].

Dans le contexte de la défense, les concepts de Bostrom sont extrêmement pertinents. Les systèmes d'IA militaire doivent être développés avec une attention particulière aux risques de comportements imprévus ou dangereux. Les stratégies d'atténuation des risques proposées par Bostrom peuvent guider les efforts pour s'assurer que les systèmes d'IA agissent de manière prévisible et sécurisée. De plus, la notion d'alignement des valeurs est cruciale pour garantir que les décisions autonomes prises par des systèmes d'IA dans des contextes de défense respectent les lois de la guerre et les droits de l'homme.

En conclusion, les travaux de Nick Bostrom sur la superintelligence offrent un cadre théorique essentiel pour comprendre les défis et les dangers potentiels des systèmes d'IA avancée. En soulignant la nécessité de stratégies robustes pour garantir la sécurité et l'alignement des IA avec les valeurs humaines, Bostrom fournit des orientations cruciales pour le développement de technologies de défense fiables et éthiques. Ses propositions encouragent une approche proactive et prudente dans la conception et le déploiement des systèmes d'IA, afin de prévenir les risques existentiels tout en maximisant les bénéfices pour la société[58].

3.Aperçu du programme Confiance.AI

L’ensemble de cette section est tiré du livre blanc sur Confiance.AI qui a été mis à la disposition du public.

3.1. Présentation Générale

Le programme Confiance.ai représente une initiative majeure de la France dans le domaine de l'intelligence artificielle (IA), lancée pour relever les défis liés à la sécurité, la certification,

[57] Bostrom, N. (2001). What is transhumanism. *Nick Bostrom*, *1998*.

[58] Bostrom, N. (2005). A history of transhumanist thought. *Journal of evolution and technology*, *14*(1).

et l'amélioration de la fiabilité des systèmes basés sur l'IA. Ce programme s'inscrit dans le contexte plus large de la stratégie nationale française sur l'IA et vise à promouvoir le développement et le déploiement de systèmes d'IA fiables dans des environnements industriels et critiques. En s'attaquant aux obstacles technologiques et en fournissant des méthodes et outils adaptés, Confiance.ai cherche à faciliter l'industrialisation de l'IA et à renforcer la compétitivité industrielle de la France dans ce secteur clé.

L'objectif principal de Confiance.ai est de développer une IA de confiance capable de s'intégrer de manière fiable et sécurisée dans des systèmes industriels et opérationnels. Les travaux d'Alan Turing ont notamment permis de s'interroger sur l'importance de la confiance dans les systèmes utilisant de l'intelligence artificielle. Pour y parvenir, le programme se concentre sur plusieurs axes de recherche et de développement, notamment l'amélioration de la robustesse et de l'explicabilité des modèles d'IA, la gestion des données, et l'évaluation de la fiabilité. Cette approche intégrée vise à créer un écosystème d'IA qui soit non seulement performant mais aussi conforme aux exigences éthiques et réglementaires.

Ainsi, Confiance.ai déploie une série de composants et outils conçus pour répondre aux défis spécifiques de l'IA dans les systèmes critiques. Parmi ceux-ci figurent des solutions pour la spécification et le workflow des données, des techniques d'apprentissage incrémental, des méthodes pour évaluer et améliorer la robustesse des modèles, ainsi que des outils pour l'explicabilité et la transparence des décisions prises par les systèmes d'IA. Ces innovations visent à renforcer la confiance des utilisateurs et des parties prenantes dans l'utilisation de l'IA pour des applications critiques.

De plus, une partie importante du programme concerne l'évaluation de la fiabilité des systèmes d'IA. Confiance.ai propose une approche unifiée qui intègre divers attributs de confiance, tels que la robustesse, la transparence, et l'éthique, dans le processus d'évaluation. Cette démarche multidimensionnelle permet d'appréhender la confiance dans l'IA de manière holistique, en considérant à la fois les aspects techniques et les enjeux éthiques et sociaux.

Malgré les progrès réalisés, le programme Confiance.ai fait face à plusieurs défis, notamment en matière de standardisation des évaluations de fiabilité et de mise en œuvre des principes éthiques dans le développement de l'IA. La collaboration avec des partenaires industriels et académiques est essentielle pour surmonter ces obstacles et pour assurer que les innovations développées soient pleinement intégrées dans les pratiques industrielles.

3.2. Implications pratiques et retombées attendues

Le programme Confiance.ai ne se limite pas à la recherche fondamentale ; il vise également à avoir des retombées concrètes et mesurables sur l'industrie française. En développant des outils et des méthodes pour intégrer l'IA de manière fiable dans les systèmes industriels, le programme cherche à répondre directement aux besoins des entreprises opérant dans des secteurs critiques tels que l'aéronautique, l'automobile, l'énergie et la santé. L'une des ambitions majeures de Confiance.ai est de faciliter l'accès des PME et des startups aux technologies d'IA de pointe, en démocratisant l'utilisation de l'IA fiable et en ouvrant de nouvelles opportunités de croissance et d'innovation pour l'ensemble du tissu industriel français.

Par son approche collaborative, réunissant des acteurs académiques, industriels et gouvernementaux, Confiance.ai fonctionne également comme un catalyseur pour l'innovation. Le programme encourage le partage des connaissances et des meilleures

pratiques, stimulant ainsi la recherche et le développement de solutions IA novatrices. Cette dynamique est essentielle pour maintenir la compétitivité de la France sur la scène internationale, dans un contexte où la maîtrise de l'IA devient un facteur clé de succès économique.

Au-delà des implications industrielles, Confiance.ai porte une attention particulière aux enjeux sociétaux et éthiques liés à l'IA. Le programme s'engage à promouvoir une IA éthique et responsable, qui respecte les droits individuels et contribue au bien commun. Cette orientation se traduit par le développement de technologies explicables et transparentes, qui permettent aux utilisateurs de comprendre et de questionner les décisions prises par les systèmes d'IA. De telles avancées sont cruciales pour renforcer la confiance du public dans l'IA et pour assurer que les bénéfices de ces technologies soient largement partagés au sein de la société.

À terme, Confiance.ai aspire à établir la France comme un leader mondial de l'IA de confiance. En mettant l'accent sur la fiabilité, la sécurité, l'éthique, et la transparence, le programme vise à définir des standards internationaux pour le développement et l'utilisation de l'IA dans les applications critiques. Cette ambition reflète la volonté de la France de jouer un rôle actif dans la définition du futur de l'IA, en proposant une vision où la technologie sert de levier pour une société plus sûre, plus juste et plus durable.

Le programme Confiance.ai illustre l'engagement de la France à relever les défis posés par l'intégration de l'IA dans les systèmes critiques, en adoptant une approche qui allie excellence scientifique, innovation industrielle, et responsabilité sociétale. En travaillant à la convergence de ces objectifs, Confiance.ai ne cherche pas seulement à développer une IA de pointe; il vise à modeler l'avenir de l'IA de manière à ce que la technologie reflète et renforce les valeurs fondamentales de notre société.

3.3. Cas d’utilisation

Le programme Confiance.ai, en abordant une diversité de cas d'utilisation exemplaires, démontre son engagement à répondre aux défis de l'intégration de l'intelligence artificielle (IA) dans des contextes industriels variés, tout en soulignant l'importance cruciale de la fiabilité et de la confiance dans les résultats produits par ces systèmes. Ces exemples illustrent l'application de l'IA dans un éventail de domaines, allant de l'automobile à l'énergie, en passant par la défense et la sécurité, et mettent en exergue la complexité et la spécificité des enjeux associés à chaque secteur.

Parmi les cas d'utilisation, l'analyse de scènes routières par Valeo pour la conduite autonome, et l'inspection visuelle de soudures chez Renault, révèlent l'importance de la précision et de la robustesse des systèmes d'IA face aux tâches critiques. L'exemple de Renault, où un modèle d'IA affiche une précision supérieure à 97% pour la détection de la qualité des soudures mais est finalement rejeté pour des raisons de fiabilité perçue, souligne l'écart entre la performance technique et la confiance des utilisateurs dans ces technologies. Ceci met en lumière le défi de construire non seulement des systèmes performants mais également des systèmes dont les opérateurs peuvent vérifier et comprendre les décisions.

Les applications dans la surveillance de l'efficacité des installations par Air Liquide et dans la détection des anomalies pour le Groupe Naval montrent comment l'IA peut transformer la maintenance préventive et la gestion des opérations en fournissant des analyses précises

basées sur des données tabulaires. Ces cas illustrent l'impact potentiel de l'IA sur l'optimisation des processus industriels et la réduction des risques opérationnels.

L'extraction d'opinions menée par Renault, basée sur l'analyse de texte, met en avant une autre dimension de l'IA : la capacité à traiter et à analyser le langage naturel pour extraire des insights précieux à partir de données non structurées. Ce cas démontre l'utilité de l'IA dans la compréhension des perceptions des consommateurs, un aspect crucial pour l'amélioration des produits et des services.

En somme, ces cas d'utilisation de Confiance.ai reflètent la volonté du programme de pousser les frontières de l'IA de confiance en développant des solutions qui allient performance technique, fiabilité, et transparence. En s'attaquant à des problèmes spécifiques dans divers secteurs, Confiance.ai contribue de manière significative à l'effort global pour aligner les avancées en IA avec les besoins industriels et les attentes sociétales, marquant ainsi une étape importante vers la réalisation d'une IA vraiment intégrée, fiable et bénéfique pour l'ensemble de la société.

4.Contexte international et comparaison avec d'autres initiatives

Dans le contexte international, l'intégration de l'intelligence artificielle (IA) dans le secteur de la défense suscite à la fois de l'enthousiasme pour ses potentielles capacités à améliorer l'efficacité des opérations et de l'inquiétude quant aux risques éthiques, sécuritaires et de contrôle. Pour cela, les travaux de Nick Bostrom sur la superintelligence semble être une piste de réflexion intéressante. Le programme français Confiance.AI se présente comme une initiative visant à instaurer un cadre de confiance pour le déploiement de l'IA dans la défense, soulignant l'importance de la transparence, de la sécurité et de l'éthique. Cette approche s'inscrit dans un mouvement plus large où divers pays et organisations internationales cherchent à établir des normes pour l'utilisation responsable de l'IA.

4.1. L’initiative américaine

Depuis plus de cinq décennies, la DARPA joue un rôle de premier plan dans la recherche et le développement (R&D) révolutionnaires, facilitant l'avancement et l'application des technologies de l'intelligence artificielle (IA) basées sur les règles et l'apprentissage statistique. Aujourd'hui, elle continue de diriger l'innovation en finançant un large éventail de programmes de R&D, allant de la recherche fondamentale au développement de technologies avancées, dans le but de réaliser l'avenir des technologies de l'IA de "Troisième Vague", qui permettront aux systèmes d'acquérir de nouvelles connaissances à travers des modèles contextuels et explicatifs génératifs[59].

En septembre 2018, la DARPA a annoncé un investissement pluriannuel de plus de 2 milliards de dollars dans de nouveaux programmes et ceux existants sous le nom de campagne "AI Next". Les domaines clés incluent l'automatisation des processus commerciaux critiques du Département de la Défense (DOD), comme la vérification des autorisations de sécurité ou l'accréditation des systèmes logiciels pour le déploiement

[59] Tadjdeh, Y. (2019). DARPA’s ‘AI next’program bearing fruit. *National Defense*, *104*(788), 8-8.

opérationnel, l'amélioration de la robustesse et de la fiabilité des systèmes d'IA, et la réduction des inefficacités en matière de puissance, de données et de performance[60].

AI Next s'appuie sur cinq décennies de création de technologies d'IA par la DARPA pour définir et façonner l'avenir, toujours en gardant à l'esprit les problèmes les plus difficiles du Département. La DARPA vise à créer des capacités puissantes pour le DOD en automatisant les processus commerciaux critiques et en avançant dans des domaines tels que l'IA robuste, l'IA adversariale, la haute performance de l'IA et la génération suivante d'algorithmes et d'applications d'IA[61].

En plus des recherches nouvelles et existantes, un composant clé de la campagne sera le programme d'Exploration de l'Intelligence Artificielle (AIE) de la DARPA, qui vise à établir la faisabilité de nouveaux concepts d'IA dans les 18 mois suivant l'attribution, avec des procédures de contractualisation et des mécanismes de financement accélérés[62].

Historiquement, l'avancement technologique a évolué les rôles des humains et des machines dans les conflits, passant de confrontations directes à des engagements médiés par des machines. La DARPA envisage un futur où les machines fonctionneront davantage comme des collègues que comme des outils, intégrant ces technologies dans des systèmes militaires collaborant avec les combattants pour faciliter de meilleures décisions dans des environnements de bataille complexes et critiques en temps. La DARPA concentre ses investissements sur une troisième vague d'IA qui apporte des machines comprenant et raisonnant dans le contexte, marquant un tournant dans l'évolution de l'intelligence artificielle vers des systèmes plus autonomes et contextuellement intelligents[63].

4.2. L’initiative chinoise

La Chine a lancé de multiples initiatives dans le domaine de l'intelligence artificielle (IA), visant à devenir un leader mondial dans ce secteur d'ici 2030. Le gouvernement chinois a publié un plan ambitieux qui souligne l'importance stratégique de l'IA pour le futur du pays. Cette vision est soutenue par d'importants investissements dans la recherche et le développement, l'éducation, et les infrastructures nécessaires pour faciliter l'innovation en IA. Par exemple, la Chine a mis en place des parcs industriels dédiés à l'IA, comme le parc d'innovation de Pékin pour l'intelligence artificielle, qui rassemble des universités, des entreprises de technologie, et des startups pour collaborer sur des projets d'IA. Ces initiatives

[60] Fouse, S., Cross, S., & Lapin, Z. (2020). DARPA’s impact on artificial intelligence. *Ai Magazine*, *41*(2), 3-8.

[61] Holzinger, A. (2021, September). The next frontier: AI we can really trust. In *Joint European conference on machine learning and knowledge discovery in databases* (pp. 427-440). Cham: Springer International Publishing.

[62] Zhang, Y., Dai, Z., Zhang, L., Wang, Z., Chen, L., & Zhou, Y. (2020, December). Application of artificial intelligence in military: From projects view. In *2020 6th International Conference on Big Data and Information Analytics (BigDIA)* (pp. 113-116). IEEE.

[63] Layton, P. (2021). Fighting Artificial Intelligence Battles: Operational Concepts for Future AI-Enabled Wars. *Network*, *4*(20), 1-100.

sont complétées par des investissements substantiels dans l'éducation et la formation en IA, visant à préparer une main-d'œuvre qualifiée pour répondre aux besoins futurs du secteur[64].

En plus de ces efforts internes, la Chine promeut également la coopération internationale en IA, participant à des forums et des conventions mondiaux pour échanger des idées et des meilleures pratiques. Le gouvernement encourage les entreprises chinoises à explorer des opportunités d'IA à l'étranger, cherchant à exporter leur technologie et leur expertise. Cela inclut des partenariats avec des entreprises et des institutions étrangères, ainsi que l'acquisition de startups innovantes dans le domaine de l'IA à travers le monde[65].

Ces initiatives ne sont pas sans controverses, notamment en ce qui concerne les implications en matière de vie privée, de sécurité des données, et d'éthique de l'IA. Le développement rapide de la technologie de reconnaissance faciale en Chine, par exemple, a soulevé des questions sur la surveillance et les droits de l'homme. Malgré cela, la Chine continue de pousser ses ambitions en IA, voyant dans cette technologie un levier crucial pour sa croissance économique et son influence globale[66].

Les efforts de la Chine dans le domaine de l'IA démontrent son engagement à être à la pointe de l'innovation technologique. Avec un soutien gouvernemental fort, une collaboration étroite entre le secteur public et privé, et une vision stratégique à long terme, la Chine se positionne comme un acteur clé dans l'évolution future de l'intelligence artificielle[67].

4.3. La réglementation européenne

Dans ses travaux, Wiener insiste sur l'importance de la régulation et de l'éthique au niveau de la cybernétique. Ainsi, l'Union européenne (UE) est à l'avant-garde de la régulation de l'intelligence artificielle (IA), cherchant à équilibrer les avantages de cette technologie avec la protection des droits fondamentaux et la sécurité des utilisateurs. En avril 2021, la Commission européenne a proposé le premier cadre juridique sur l'IA, connu sous le nom de règlement sur l'intelligence artificielle (AI Act), marquant une étape significative vers la mise en place d'une réglementation harmonisée au sein de l'UE. Ce règlement vise à instaurer un écosystème de confiance autour de l'IA, en classant les systèmes d'IA selon quatre niveaux de risque : inacceptable, élevé, limité, et minimal. Les applications d'IA considérées comme présentant un risque inacceptable seront interdites, tandis que celles à risque élevé seront soumises à des exigences strictes en matière de transparence, de sécurité et de supervision[68].

Le cadre réglementaire européen se distingue par son approche fondée sur les risques et son objectif de garantir la sécurité, la transparence et le respect des droits de l'homme, cela fait, entre autres, écho aux idées développées par Nick Bostrom. Par exemple, il exige des

[64] Ma, A. (2018). L'intelligence artificielle en Chine: un état des lieux.

[65] Fischer, S. C. (2018). Intelligence artificielle: les ambitions de la Chine. *Politique de sécurité: analyses du CSS*, *220*.

[66] Thibout, C. (2019). La compétition mondiale de l'intelligence artificielle. *Pouvoirs*, (3), 131-142.
[67] Cugurullo, F. (2021). «One AI to rule them all». En Chine, l'unification de la gouvernance urbaine par l'intelligence artificielle. *GREEN*, (1), 134-137.

[68] Petel, A. (2022). Quelle réglementation européenne sur l'intelligence artificielle?. *I2D–Information, données & documents*, (1), 22-28.

évaluations d'impact sur la protection des données et la conformité aux normes éthiques avant le déploiement de certaines applications d'IA. En outre, il promeut l'innovation responsable en encourageant le développement d'une IA fiable et sûre[69].

Cette initiative de l'UE a été largement saluée comme un pas en avant vers la création d'un cadre réglementaire mondial pour l'IA, incitant d'autres régions à envisager des législations similaires. Cependant, le processus législatif européen étant complexe et impliquant de nombreux acteurs, l'adoption finale du règlement sur l'IA pourrait prendre plusieurs années. Une fois adopté, ce règlement pourrait servir de modèle pour la régulation de l'IA à l'échelle mondiale, établissant des normes élevées en matière d'éthique et de sûreté[70].

L'approche réglementaire de l'UE en matière d'IA reflète son engagement à protéger les citoyens tout en favorisant l'innovation et la compétitivité de son économie numérique. Elle souligne l'importance d'une gouvernance de l'IA qui soit à la fois éthique, sûre et bénéfique pour la société.

Par conséquent, nous avons pu analyser qu'aux États-Unis, le Département de la Défense a publié en 2018 une stratégie d'IA qui souligne l'importance de développer une IA éthique, sûre et efficace pour maintenir l'avantage stratégique américain. Cette stratégie est complétée par la création du Joint Artificial Intelligence Center (JAIC), qui vise à accélérer l'adoption de l'IA à travers les différentes branches des forces armées américaines[71].

En Chine, le gouvernement a intégré l'IA dans sa stratégie de modernisation militaire, visant à devenir un leader mondial en matière de technologies d'IA d'ici 2030. Bien que les détails spécifiques de l'approche chinoise en matière d'éthique et de sûreté de l'IA dans la défense soient moins transparents, l'accent est mis sur le développement rapide et l'application de l'IA pour renforcer les capacités militaires.

L'Union européenne, quant à elle, promeut une approche coordonnée pour développer une IA de confiance, comme le reflète le futur règlement européen sur l'IA. Bien que centrée sur les applications civiles, cette législation pourrait influencer les initiatives de défense des États membres, en mettant l'accent sur la sécurité, l'éthique et la transparence.

La comparaison de "Confiance.AI" avec d'autres initiatives internationales révèle une diversité d'approches, chacune reflétant les priorités stratégiques, les cadres réglementaires et les cultures de sécurité nationales. Cependant, un dénominateur commun émerge : la reconnaissance de l'importance cruciale de l'IA pour l'avenir de la défense et la nécessité d'un cadre de confiance pour en maximiser les bénéfices tout en minimisant les risques.

4.4 Différences d'approches entre la France et les États-Unis en matière d'IA

[69] Martin, A. (2022, June). L'intelligence artificielle peut-elle être saisie par le droit de l'Union européenne?. In *Conférence Nationale en Intelligence Artificielle 2022 (CNIA 2022)*.

[70] Castillo, M. (2023). L'Union européenne: vers la maîtrise de l'intelligence artificielle?. *Cahiers de la recherche sur les droits fondamentaux*, (21), 99-107.

[71] Bastian, N. D. (2020). Building the Army's Artificial Intelligence Workforce. *The Cyber Defense Review*, *5*(2), 59-64.

Les différences d'approches entre la France et les États-Unis dans le développement et la régulation de l'intelligence artificielle (IA) sont significatives, influencées par des contextes culturels, économiques et politiques distincts.

En matière d'innovation et de régulation, les États-Unis privilégient une approche axée sur l'innovation et la flexibilité, favorisant un environnement où les startups technologiques et les grandes entreprises bénéficient d'une régulation relativement légère. Cette approche est conçue pour encourager l'innovation rapide et l'adaptation aux évolutions technologiques. Par exemple, des organismes comme la Federal Trade Commission (FTC)[72] et le National Institute of Standards and Technology (NIST)[73] se concentrent sur l'élaboration de cadres flexibles et sur l'autorégulation des entreprises, permettant ainsi une croissance rapide et une innovation continue sans lourdeur réglementaire excessive.

En revanche, la France adopte une approche plus régulatrice et éthique, fortement influencée par l'Union européenne[74]. La régulation française de l'IA met l'accent sur la protection des droits individuels et la mise en place de normes strictes pour garantir une utilisation responsable de l'IA[75]. Le programme Confiance.ai illustre cette orientation, en cherchant à développer des systèmes d'IA qui sont non seulement performants mais aussi sécurisés, transparents et éthiques[76]. Par exemple, les lignes directrices de la Commission Européenne pour une IA digne de confiance insistent sur l'importance de la transparence, de l'équité et de la responsabilité dans le développement de l'IA.

Aux États-Unis, la collaboration entre le secteur public et privé est souvent initiée par des entreprises privées avec des partenariats ponctuels avec des agences gouvernementales. Par exemple, des entreprises comme Google et Microsoft collaborent avec des agences comme la Defense Advanced Research Projects Agency (DARPA)[77] sur des projets spécifiques, tout en conservant une grande autonomie dans leurs recherches et développements.

En France, l'approche est plus centralisée et structurée, avec des initiatives nationales telles que Confiance.ai qui impliquent une collaboration étroite entre le gouvernement, les entreprises et les institutions de recherche. Ce modèle favorise une coordination des efforts et des ressources, visant à aligner les objectifs industriels et les politiques publiques. Par exemple, Confiance.ai rassemble des acteurs industriels comme Renault, Valeo et Thales,

[72] Federal Trade Commission. (2020). Using Artificial Intelligence and Algorithms.

[73] National Institute of Standards and Technology. (2020). AI and Machine Learning.

[74] Commission Européenne. (2020). Lignes directrices pour une IA digne de confiance.

[75] Ministère de l'Économie, des Finances et de la Relance. (2021). Stratégie nationale pour l'intelligence artificielle.

[76] Confiance.ai. (2021). Programme de développement et d'intégration de l'IA.

[77] Defense Advanced Research Projects Agency. (2020). Artificial Intelligence Explorations.

ainsi que des institutions de recherche comme l'INRIA[78], pour développer des technologies d'IA fiables et éthiques, adaptées aux besoins industriels français.

En résumé, les États-Unis et la France adoptent des approches distinctes en matière de développement et de régulation de l'IA. Les États-Unis misent sur l'innovation rapide et l'autorégulation pour maintenir leur leadership technologique, tandis que la France, soutenue par l'Union européenne, privilégie une régulation stricte et une approche collaborative pour garantir que les systèmes d'IA soient sécurisés, transparents et éthiques. Ces différences reflètent des priorités et des valeurs sociétales variées, influençant la manière dont chaque pays intègre l'IA dans ses infrastructures industrielles et sociales.

4.5. Différences d'approches entre la France et la Chine en matière d'IA

La France et la Chine adoptent des approches fondamentalement différentes en matière de développement et de régulation de l'intelligence artificielle (IA), influencées par leurs structures politiques, leurs priorités stratégiques et leurs cadres de gouvernance.

La Chine adopte une approche centralisée et dirigiste dans le développement de l'IA, caractérisée par une forte intervention de l'État. Le gouvernement chinois joue un rôle prédominant dans la planification stratégique et l'investissement massif dans l'IA. Le plan "Next Generation Artificial Intelligence Development Plan" de 2017[79] vise à faire de la Chine le leader mondial de l'IA d'ici 2030, avec des objectifs clairs et des ressources substantielles allouées pour atteindre ces objectifs. Cette centralisation se traduit par un contrôle étroit de l'État sur les entreprises technologiques et l'utilisation des données, souvent au détriment des considérations de vie privée et des droits individuels.

En France, l'approche est plus équilibrée, mettant l'accent sur l'innovation tout en respectant les normes éthiques et les droits individuels. La régulation française, fortement influencée par l'Union européenne, vise à protéger les droits des citoyens tout en promouvant une IA éthique et transparente. Le programme Confiance.ai en est un exemple, cherchant à développer des systèmes d'IA sécurisés et responsables, avec une attention particulière à la transparence et à l’explicabilité[80].

Les priorités stratégiques de la Chine en matière d'IA incluent une large utilisation de l'IA pour le contrôle social et la surveillance. Les technologies d'IA sont intégrées dans les systèmes de sécurité publique pour la surveillance de masse, le contrôle social et la gestion des populations. Cette approche soulève des préoccupations éthiques internationales, notamment en ce qui concerne la protection des libertés civiles et la vie privée.

En France, les priorités sont orientées vers des applications de l'IA qui respectent les normes éthiques et les valeurs démocratiques. Le pays se concentre sur des secteurs tels que la santé, l'énergie et l'industrie, où l'IA peut apporter des améliorations significatives tout en étant

[78] INRIA. (2021). Partenariat pour une IA éthique et fiable.

[79] The State Council of the People's Republic of China. (2017). Next Generation Artificial Intelligence Development Plan.

[80] Commission Européenne. (2020). Lignes directrices pour une IA digne de confiance.

utilisée de manière responsable et éthique. La France met un accent particulier sur la transparence et l'explicabilité des systèmes d'IA, afin de renforcer la confiance des utilisateurs et du public[81].

La régulation de l'IA en Chine est largement axée sur la réalisation des objectifs stratégiques nationaux, avec une flexibilité pour les entreprises tant qu'elles alignent leurs efforts sur les directives de l'État. Le contrôle centralisé permet une mise en œuvre rapide des technologies d'IA, mais au prix d'une surveillance stricte et d'un contrôle étatique accru[82].

En France, la régulation de l'IA est fortement influencée par les lignes directrices de l'Union européenne pour une IA digne de confiance, qui insistent sur l'éthique, la transparence et la protection des droits individuels. Les initiatives comme Confiance.ai visent à intégrer ces principes dès les premières étapes du développement technologique, garantissant que l'IA est utilisée de manière responsable et bénéfique pour la société.

Les approches de la France et de la Chine en matière de développement et de régulation de l'IA reflètent des différences fondamentales en termes de valeurs et de priorités. La Chine mise sur un contrôle étatique centralisé pour accélérer son leadership en IA, souvent au détriment des droits individuels, tandis que la France privilégie une approche équilibrée qui combine innovation technologique et respect des normes éthiques et des droits humains. Ces différences ont des implications profondes sur la manière dont chaque pays intègre l'IA dans ses infrastructures industrielles et sociétales.

Pour conclure, les approches françaises, chinoises et américaines en matière de développement et de régulation de l'intelligence artificielle (IA) reflètent des différences profondes enracinées dans leurs contextes culturels, économiques et politiques.

Les États-Unis privilégient une approche axée sur l'innovation et la flexibilité, soutenue par un écosystème dynamique de startups, de grandes entreprises technologiques et de financements privés massifs. La régulation y est généralement plus légère, avec un fort accent sur l'autorégulation des entreprises et l'élaboration de cadres flexibles pour encourager l'innovation rapide. Cette approche favorise une croissance rapide et une adaptation continue aux évolutions technologiques, mais peut manquer de garde-fous stricts pour la protection des droits individuels et l'éthique.

La Chine, quant à elle, adopte une approche centralisée et dirigiste, caractérisée par une forte intervention de l'État. Le gouvernement chinois joue un rôle prédominant dans la planification stratégique et l'investissement massif dans l'IA, visant à faire du pays le leader mondial de l'IA. Cette approche se traduit par un contrôle étroit de l'État sur les entreprises technologiques et l'utilisation des données, souvent au détriment des considérations de vie privée et des droits individuels. Les priorités stratégiques incluent l'utilisation de l'IA pour le contrôle social et la surveillance de masse, soulevant des préoccupations éthiques internationales.

Enfin la France, influencée par l'Union européenne, adopte une approche équilibrée entre innovation technologique et régulation stricte. La régulation française met l'accent sur la

[81] Confiance.ai. (2021). Programme de développement et d'intégration de l'IA.

[82] Xinhua News Agency. (2018). China Focus: China's AI Development Plan Takes Shape.

protection des droits individuels et l'éthique, avec des initiatives comme Confiance.ai visant à développer des systèmes d'IA sécurisés, transparents et éthiques. La France favorise également une collaboration étroite entre le secteur public et privé, avec une coordination centralisée des efforts pour aligner les objectifs industriels et les politiques publiques.

Pour se démarquer des autres pays européens tout en respectant la législation européenne en matière d'intelligence artificielle, la France peut capitaliser sur plusieurs aspects uniques de son approche :

Tout d'abord, la France peut renforcer son rôle de leader dans la régulation éthique de l'IA en s'assurant que ses normes et pratiques non seulement respectent mais surpassent les lignes directrices de l'Union européenne. En proposant des cadres de régulation qui intègrent profondément les principes de transparence, d'explicabilité et de protection des droits individuels, la France peut établir des standards élevés pour le reste de l'Europe.

En promouvant une innovation responsable, la France peut se positionner comme un pionnier dans le développement de technologies d'IA qui équilibrent performance et éthique. Le programme Confiance.ai est un exemple de cette approche, en développant des systèmes d'IA robustes et fiables qui peuvent servir de modèle pour d'autres pays européens.

De plus, elle peut également jouer un rôle clé en facilitant des partenariats stratégiques entre les acteurs publics et privés, ainsi qu'entre les pays européens. En renforçant les collaborations transfrontalières, la France peut aider à créer un écosystème d'IA européen cohérent et intégré, capable de rivaliser avec les puissances technologiques mondiales.

Pour finir, en investissant dans la formation et l'éducation, la France peut développer une expertise nationale de haut niveau en IA. En créant des programmes éducatifs robustes et en soutenant la recherche académique, la France peut s'assurer que ses talents en IA sont parmi les meilleurs au monde, ce qui est crucial pour maintenir une compétitivité à long terme.

En combinant ces éléments, la France peut non seulement se démarquer au sein de l'Europe mais aussi contribuer de manière significative à l'élaboration d'un cadre européen pour une IA éthique et responsable, tout en respectant les législations européennes existantes. Cette approche équilibrée entre innovation, régulation et collaboration pourrait positionner la France comme un leader mondial dans le domaine de l'intelligence artificielle.

En effet, l'approche de la France et, plus largement, de l'Union Européenne (UE) en matière d'intelligence artificielle (IA) se distingue par un fort engagement envers les principes éthiques et la protection des droits fondamentaux. Cette orientation éthique est ancrée dans une série de lignes directrices et de régulations conçues pour assurer que le développement et l'utilisation de l'IA respectent des normes élevées de transparence, de responsabilité et de justice. Par exemple, les "Lignes directrices pour une IA digne de confiance" publiées par la Commission Européenne insistent sur la nécessité d'intégrer les principes de respect de la dignité humaine, de prévention des préjudices, d'équité, de transparence et de responsabilité dès les premières étapes du développement technologique. Ces principes se traduisent par des exigences concrètes telles que l'obligation de garantir l'explicabilité des décisions prises par les systèmes d'IA, permettant ainsi aux utilisateurs de comprendre et de contester ces décisions si nécessaire[83].

En France, le programme Confiance.ai incarne cet engagement éthique en cherchant à développer des technologies d'IA qui sont non seulement performantes mais aussi sécurisées

[83] Commission Européenne. (2019). Lignes directrices pour une IA digne de confiance.

et transparentes. Ce programme vise à créer un cadre de confiance autour de l'IA en garantissant que les systèmes respectent les normes éthiques et les droits individuels, tout en répondant aux besoins industriels et sociétaux. En parallèle, l'UE promeut des initiatives comme la création de comités éthiques indépendants pour surveiller le développement de l'IA et l'établissement de cadres de certification pour les technologies d'IA, assurant ainsi une supervision rigoureuse et continue[84].

Cette approche éthique vise à instaurer une confiance durable dans les technologies d'IA, en veillant à ce que celles-ci contribuent positivement à la société sans compromettre les valeurs fondamentales de l'humanité. En mettant l'accent sur l'équilibre entre innovation et protection des droits, la France et l'UE se positionnent comme des leaders mondiaux dans la promotion d'une IA responsable et bénéfique pour tous.

4.6. Risques globaux encourus par l'Intelligence Artificielle

L'introduction de l'intelligence artificielle (IA) dans divers aspects de la société soulève de nombreux risques éthiques qui nécessitent une attention particulière. L'un des principaux risques est celui du biais algorithmique. Les systèmes d'IA apprennent à partir de données existantes, et si ces données sont biaisées, les décisions prises par l'IA peuvent perpétuer ou même amplifier ces biais, conduisant à des discriminations injustes. Par exemple, des systèmes de reconnaissance faciale ont été critiqués pour leur manque de précision dans l'identification des personnes de couleur par rapport aux personnes blanches, ce qui peut entraîner des discriminations systémiques dans l'application de la loi[85].

Un autre risque majeur est lié à la transparence et à l'explicabilité des systèmes d'IA. Les décisions prises par des algorithmes complexes sont souvent opaques et difficiles à comprendre pour les utilisateurs finaux, ce qui pose des problèmes de responsabilité et de confiance. Si une décision prise par un système d'IA affecte négativement une personne, il peut être difficile de déterminer comment et pourquoi cette décision a été prise, rendant difficile la contestation ou la correction des erreurs[86].

La vie privée est également un enjeu crucial. Les systèmes d'IA reposent souvent sur la collecte et l'analyse de grandes quantités de données personnelles, soulevant des préoccupations concernant la surveillance et le respect de la vie privée des individus. Sans des cadres robustes de protection des données, il existe un risque que les informations personnelles soient utilisées de manière abusive ou sans le consentement des individus[87].

En outre, le développement d'armes autonomes soulève des questions éthiques graves concernant la délégation de décisions de vie ou de mort à des machines. L'absence de

[84] High-Level Expert Group on AI. (2019). Ethics Guidelines for Trustworthy AI.

[85] Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency.

[86] Doshi-Velez, F., & Kim, B. (2017). Towards a Rigorous Science of Interpretable Machine Learning. arXiv preprint arXiv:1702.08608.

[87] Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs.

contrôle humain direct sur ces systèmes pourrait mener à des violations du droit international humanitaire et à des conséquences imprévisibles sur le champ de bataille[88].

Enfin, la perte d'emplois due à l'automatisation par l'IA représente un défi éthique et socio-économique. Si l'IA remplace un grand nombre de travailleurs humains, cela pourrait entraîner des inégalités économiques accrues et des tensions sociales, nécessitant des politiques pour gérer la transition et protéger les travailleurs affectés[89].

En somme, bien que l'IA offre de nombreuses opportunités pour améliorer divers aspects de la vie humaine, elle comporte également des risques éthiques significatifs qui doivent être soigneusement gérés pour assurer un déploiement responsable et équitable de ces technologies.

III. Résultats et Implications

1. Compréhension des enjeux liés à l'IA dans la défense

La compréhension des enjeux liés à l'intelligence artificielle (IA) dans le secteur de la défense est cruciale pour le développement de stratégies militaires et sécuritaires efficaces à l'ère numérique. L'intégration de l'IA dans les systèmes de défense présente des opportunités significatives, telles que l'amélioration de la précision et de l'efficacité des opérations militaires, la gestion des grandes quantités de données pour le renseignement, et le développement de systèmes autonomes pour des missions de reconnaissance et de combat. Cependant, ces avancées technologiques introduisent également des défis complexes sur les plans éthique, sécuritaire et juridique.

1.1 Le volet sécuritaire

L'intégration de l'intelligence artificielle (IA) dans les systèmes de défense pose des enjeux sécuritaires significatifs, nécessitant une attention particulière aux vulnérabilités potentielles et aux mesures de mitigation des risques. Ces aspects sont abordés dans les réalisations de Nick Bostrom. L'IA peut considérablement améliorer les capacités de défense en permettant une analyse rapide et précise des grandes quantités de données pour le renseignement et la surveillance, en optimisant la prise de décision opérationnelle et en automatisant certaines tâches critiques. Par exemple, les systèmes d'IA peuvent être utilisés pour détecter des anomalies dans les comportements ou les mouvements des troupes adverses, fournissant ainsi des alertes précoces et améliorant la réactivité des forces armées[90].

[88] Scharre, P. (2018). Army of None: Autonomous Weapons and the Future of War. W. W. Norton & Company.

[89] Brynjolfsson, E., & McAfee, A. (2014). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. W. W. Norton & Company.

[90] Cummings, M. L. (2017). Artificial Intelligence and the Future of Warfare. Chatham House.

Cependant, cette dépendance accrue aux systèmes d'IA expose également les infrastructures de défense à des cyberattaques sophistiquées. Les adversaires peuvent cibler les algorithmes d'IA, soit par des attaques adverses qui manipulent les données d'entrée pour provoquer des erreurs dans les prédictions, soit par des intrusions visant à corrompre les modèles d'IA ou à voler des informations sensibles[91]. La vulnérabilité des systèmes d'IA à ces attaques peut compromettre la sécurité nationale, en perturbant les opérations militaires et en fournissant de fausses informations stratégiques.

Pour faire face à ces défis, il est crucial de développer des systèmes d'IA robustes et résilients. Cela inclut la mise en place de mesures de cybersécurité avancées, telles que la détection et la prévention des intrusions, l'audit continu des algorithmes d'IA et la formation des modèles sur des jeux de données diversifiés pour améliorer leur robustesse face aux attaques adverses. Le rapport de l'OTAN sur les implications de l'IA pour la cybersécurité souligne la nécessité d'une approche proactive en matière de sécurité, en intégrant la résilience dès la conception des systèmes d'IA et en assurant une surveillance continue de leurs performances et de leurs vulnérabilités[92].

En conclusion, le volet sécuritaire de l'intégration de l'IA dans la défense est double : il s'agit d'exploiter les capacités avancées de l'IA pour renforcer la défense tout en assurant une protection rigoureuse contre les cybermenaces. Une attention continue à la robustesse des systèmes et à la mitigation des risques est essentielle pour garantir que les avantages de l'IA soient pleinement réalisés sans compromettre la sécurité nationale.

1.2 Le volet éthique

L'utilisation de l'intelligence artificielle (IA) dans le domaine de la défense soulève des questions éthiques profondes et complexes, notamment en ce qui concerne l'autonomie des systèmes d'armes létaux. L'un des principaux débats éthiques porte sur la délégation de décisions de vie ou de mort à des machines, ce qui pose des problèmes de responsabilité morale et juridique[93]. Les systèmes d'armes autonomes, tels que les drones armés ou les robots de combat, peuvent agir sans intervention humaine directe, ce qui soulève des inquiétudes quant à la capacité de ces systèmes à respecter les principes du droit international humanitaire, tels que la distinction, la proportionnalité et la nécessité militaire[94].

---

[91] Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. arXiv preprint arXiv:1802.07228.

[92] NATO. (2020). Science & Technology Trends 2020-2040: Exploring the S&T Edge.

[93] Crootof, R. (2015). The Killer Robots Are Here: Legal and Policy Implications. Cardozo Law Review, 36(1), 183-201.

[94] Binnendijk, A. (2020). The Role of Artificial Intelligence in Future Warfare. RAND Corporation.

De plus, l'absence de transparence et d'explicabilité des algorithmes d'IA aggrave les préoccupations éthiques. Les décisions prises par des systèmes d'IA peuvent être opaques et difficiles à comprendre, même pour leurs concepteurs, ce qui complique la détermination de la responsabilité en cas de faute ou de violation des droits humains. Par ailleurs, l'utilisation de l'IA dans la défense peut exacerber les inégalités mondiales, les pays technologiquement avancés ayant un accès disproportionné aux capacités de l'IA, ce qui pourrait renforcer leur supériorité militaire et accroître les tensions internationales[95].

Les chercheurs et les responsables politiques appellent à une régulation stricte et à des cadres éthiques robustes pour encadrer l'utilisation de l'IA dans le domaine militaire. Des initiatives comme le Partenariat mondial sur l'intelligence artificielle (GPAI) et les discussions au sein des Nations Unies sur les systèmes d'armes létaux autonomes (SALA) cherchent à établir des normes internationales pour garantir que le développement et l'utilisation de l'IA respectent les valeurs éthiques fondamentales et les droits de l'homme. En conclusion, l'intégration de l'IA dans la défense nécessite une vigilance éthique constante pour équilibrer les avantages technologiques avec la protection des principes humanitaires et la stabilité mondiale[96].

1.3. Le volet juridique

L'intégration de l'intelligence artificielle (IA) dans les systèmes de défense pose des défis juridiques significatifs, notamment en ce qui concerne la conformité avec le droit international humanitaire (DIH) et le droit des conflits armés. L'un des principaux enjeux juridiques concerne la responsabilité en cas de violation des lois de la guerre par des systèmes d'armes autonomes. Selon le DIH, les principes de distinction, de proportionnalité et de précaution doivent être respectés lors des opérations militaires. Cependant, l'autonomie des systèmes d'IA rend difficile l'attribution de la responsabilité légale, notamment lorsqu'il s'agit de déterminer qui – le concepteur, le programmeur, l'opérateur ou l'État – doit être tenu responsable en cas de dommages collatéraux ou de violations des droits humains[97].

En outre, les systèmes d'armes autonomes soulèvent des questions juridiques concernant la capacité des machines à respecter et à interpréter les règles complexes du DIH. Par exemple, la Convention de Genève impose des obligations strictes concernant le traitement des civils et des prisonniers de guerre, des obligations qui nécessitent souvent des jugements contextuels et

---

[95] Horowitz, M. C. (2019). The Ethics & Morality of Robotic Warfare: Assessing the Debate over Autonomous Weapons. Daedalus, 148(2), 25-36.

[96] UNIDIR. (2017). The Weaponization of Increasingly Autonomous Technologies: Considering Ethics and Social Values. United Nations Institute for Disarmament Research.

[97] Schmitt, M. N., & Thurnher, J. S. (2013). "Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict. Harvard National Security Journal, 4, 231-281.

éthiques que les algorithmes d'IA ne sont pas encore capables de faire de manière fiable. La question de l'accountability (responsabilité) juridique est également centrale, car l'opacité des algorithmes et la difficulté à retracer les décisions automatisées compliquent l'application de sanctions légales et la réparation des victimes[98].

Les initiatives internationales, telles que les discussions au sein des Nations Unies sur les systèmes d'armes létaux autonomes (SALA), cherchent à créer des cadres juridiques adaptés à ces nouvelles technologies. Ces discussions visent à combler les lacunes actuelles du droit international et à garantir que les innovations technologiques respectent les obligations légales et les droits humains. Par exemple, le Groupe d'experts gouvernementaux de l'ONU travaille sur la clarification et le développement de normes pour l'utilisation des technologies autonomes dans les conflits armés. En conclusion, l'utilisation de l'IA dans la défense exige une adaptation et un renforcement des cadres juridiques existants pour garantir que les nouvelles technologies soient déployées de manière responsable et conforme au droit international[99].

En résumé, la compréhension des enjeux liés à l'IA dans la défense est essentielle pour naviguer dans les opportunités et les défis que ces technologies présentent. Il est impératif de développer des systèmes d'IA qui sont non seulement efficaces et innovants, mais aussi sécurisés, éthiques et conformes aux normes légales. La collaboration internationale, la recherche continue et le dialogue entre les secteurs public et privé sont essentiels pour garantir une intégration responsable et bénéfique de l'IA dans la défense.

2.Évaluation de l'impact et de l'efficacité de Confiance.AI

L'évaluation de l'impact et de l'efficacité du programme français Confiance.AI montre des résultats prometteurs tout en révélant des domaines nécessitant des améliorations continues pour garantir son succès à long terme. Depuis son lancement, Confiance.AI a joué un rôle crucial dans la promotion d'une IA éthique et fiable en France, contribuant à positionner le pays comme un acteur dans le domaine de l'intelligence artificielle.

L'un des succès notables du programme est la mise en œuvre de projets pilotes dans divers secteurs critiques, tels que la santé, les transports et l'énergie. Dans le domaine de la santé, par exemple, Confiance.AI a permis le développement de systèmes de diagnostic assistés par IA qui ont amélioré la précision et la rapidité des diagnostics médicaux. Ces systèmes utilisent des algorithmes sophistiqués pour analyser des images médicales et d'autres données de patients, aidant ainsi les médecins à identifier des conditions médicales complexes plus rapidement et avec une plus grande précision. De plus, ces technologies respectent strictement les règles de confidentialité des données, assurant ainsi la protection des informations personnelles des patients. Cette avancée a non seulement amélioré les soins aux patients, mais a également

[98] Boulanin, V., & Verbruggen, M. (2017). Mapping the Development of Autonomy in Weapon Systems. SIPRI.

[99] United Nations Office for Disarmament Affairs (UNODA). (2019). Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapon Systems.

démontré la capacité de l'IA à transformer les pratiques médicales tout en respectant les normes éthiques[100].

Dans le secteur des transports, Confiance.AI a soutenu le développement de véhicules autonomes et de systèmes de gestion du trafic intelligents. Ces innovations ont le potentiel de réduire les accidents de la route, de diminuer la congestion urbaine et de rendre les transports publics plus efficaces. Les projets soutenus par le programme ont montré que l'IA peut être utilisée pour optimiser les itinéraires, prévoir les pannes de véhicules et améliorer globalement la sécurité routière. Par exemple, des algorithmes de machine learning peuvent analyser des données en temps réel provenant de capteurs et de caméras pour prendre des décisions rapides et adaptées, contribuant ainsi à une meilleure fluidité du trafic et à une réduction des émissions polluantes[101].

En outre, Confiance.AI a renforcé la collaboration entre les acteurs publics et privés, créant un écosystème d'innovation robuste. Les partenariats avec des entreprises technologiques, des instituts de recherche et des universités ont conduit à des avancées significatives dans la recherche et le développement de solutions d'IA sécurisées et éthiques. Par exemple, des collaborations entre des start-ups innovantes et des grands groupes industriels ont permis de développer des technologies d'IA avancées qui sont maintenant utilisées dans divers secteurs. Cette synergie entre le public et le privé a également favorisé le transfert de connaissances et le partage des meilleures pratiques, accélérant ainsi le développement et l'adoption de l'IA en France[102].

Un autre aspect crucial du programme Confiance.AI est l'accent mis sur la formation et la sensibilisation. Le programme a mis en place de nombreuses initiatives pour former les professionnels aux enjeux éthiques et techniques de l'IA. Des ateliers, des cours en ligne et des programmes de certification ont été développés pour aider les travailleurs à acquérir les compétences nécessaires pour travailler avec les technologies de l'IA . Ces initiatives de formation ont contribué à la montée en compétence de la main-d'œuvre française, préparant ainsi le pays à relever les défis et à saisir les opportunités offertes par l'IA. Par exemple, des programmes de formation spécifiques ont été mis en place pour les professionnels de la santé, les ingénieurs en informatique et les décideurs politiques, leur permettant de mieux comprendre et de mieux intégrer l'IA dans leurs pratiques quotidiennes[103].

Cependant, malgré ces succès, l'évaluation de Confiance.AI a également identifié plusieurs défis. L'un des principaux obstacles est la nécessité de mettre en place des régulations efficaces

[100] Secrétariat général pour l'investissement (SGPI). (2021). Bilan et perspectives du programme Confiance.AI. Rapport officiel.

[101] AI for Humanity. (2020). Évaluation du programme Confiance.AI. Ministère de l'Économie et des Finances.

[102] Institut Montaigne. (2021). Réussir l'Intelligence Artificielle: Bilan et propositions pour la France. Rapport de l'Institut Montaigne.

[103] Villani, C. (2018). Donner un sens à l'intelligence artificielle: Pour une stratégie nationale et européenne. Rapport au Premier ministre.

et adaptatives pour suivre le rythme rapide des innovations technologiques. Les régulateurs doivent continuellement adapter les cadres législatifs et réglementaires pour s'assurer qu'ils restent pertinents face aux avancées de l'IA. Par exemple, des régulations spécifiques concernant les véhicules autonomes, les drones et les systèmes d'IA dans la santé doivent être élaborées pour garantir leur utilisation sûre et éthique. Cela inclut la définition de normes de sécurité, la mise en place de protocoles de test rigoureux et l'établissement de mécanismes de responsabilité clairs en cas de défaillance[104].

Un autre défi identifié est la nécessité de renforcer les mécanismes de gouvernance pour superviser l'application des principes éthiques dans les projets d'IA. Les initiatives de gouvernance doivent garantir que les projets d'IA respectent les valeurs éthiques fondamentales et protègent les droits de l'homme. Cela peut inclure la création de comités d'éthique indépendants pour évaluer les projets d'IA, l'établissement de lignes directrices claires pour l'utilisation responsable de l'IA et la mise en place de systèmes de surveillance pour détecter et prévenir les abus. Par exemple, des comités d'éthique pourraient être chargés d'examiner les projets de recherche en IA pour s'assurer qu'ils respectent les normes éthiques et ne posent pas de risques pour les individus ou la société[105].

En outre, l'évaluation de l'impact de Confiance.AI souligne l'importance d'une approche inclusive et participative. Impliquer un large éventail de parties prenantes, y compris les citoyens, les organisations de la société civile et les experts en éthique, dans le développement et la mise en œuvre des projets d'IA est essentiel pour garantir que les technologies développées répondent aux besoins et aux préoccupations de la société. Par exemple, des consultations publiques et des ateliers participatifs pourraient être organisés pour recueillir les avis et les suggestions des citoyens sur les projets d'IA, assurant ainsi une plus grande transparence et une meilleure acceptation des technologies par le public[106].

Enfin, l'évaluation recommande de renforcer la coopération internationale. Face à la nature mondiale des technologies de l'IA, une collaboration accrue avec d'autres pays et organisations internationales est nécessaire pour harmoniser les normes et partager les meilleures pratiques. Des partenariats internationaux peuvent aider à relever les défis communs, tels que la cybersécurité, la protection des données et l'éthique de l'IA. Par exemple, la France pourrait jouer un rôle de premier plan dans les discussions internationales sur la régulation de l'IA, en collaborant avec des organisations comme l'Union européenne, l'Organisation des Nations Unies et l'Organisation de coopération et de développement économiques (OCDE).

En conclusion, bien que le programme Confiance.AI ait eu un impact positif sur le développement de l'IA en France, son efficacité à long terme dépendra de la capacité à surmonter les défis réglementaires, éthiques et de gouvernance. Maintenir un engagement fort

[104] European Commission. (2020). White Paper on Artificial Intelligence: A European approach to excellence and trust. European Commission.

[105] OECD. (2021). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments.

[106] UNIDIR. (2017). The Weaponization of Increasingly Autonomous Technologies: Considering Ethics and Social Values. United Nations Institute for Disarmament Research.

envers les valeurs éthiques, renforcer les mécanismes de régulation et de gouvernance, et promouvoir une coopération internationale active seront essentiels pour garantir que les technologies de l'IA développées et déployées respectent les principes de confiance, de sécurité et de respect des droits humains[107].

3.Comparaison avec les initiatives internationales pour une IA de confiance

L'intelligence artificielle (IA) révolutionne de nombreux aspects de nos vies, mais elle pose également des défis en termes de fiabilité, d'éthique et d'équité. Pour répondre à ces enjeux, plusieurs initiatives internationales collaborent pour développer une IA de confiance. Voici un tour d'horizon de ces initiatives, en lien avec le programme français Confiance.ai et poursuivant les interrogations qu'avaient pu soulever les travaux d'Alan Turing

Confiance.IA au Canada (Québec)

Soutenu par le CRIM, Confiance.IA est un consortium québécois réunissant acteurs privés et publics pour développer des méthodes et outils pour une IA durable, éthique, sûre et responsable. Ce programme cible notamment les secteurs réglementés comme l'aéronautique, les transports et la finance. La collaboration entre Confiance.ai et Confiance.IA se renforce, facilitant l'échange de savoir-faire et de pratiques[108].

ZERTIFIZIERTE KI en Allemagne

Dirigé par le Fraunhofer IAIS, l'Institut allemand de normalisation (DIN) et l'Office fédéral allemand de la sécurité de l'information (BSI), ce projet vise à développer des procédures de test pour l'évaluation des systèmes d'IA, garantissant leur fiabilité technique et une utilisation responsable. Il intègre les besoins industriels et se concentre sur le développement de critères d'évaluation, de mesures de sauvegarde, de standards d'IA, et l'exploration de nouveaux modèles économiques. Confiance.ai collabore étroitement avec ZERTIFIZIERTE KI, notamment à travers des événements comme le symposium AITA[109].

VDE en Allemagne

VDE est une organisation technologique européenne de premier plan, connue pour ses normes de sécurité et son rôle en matière de normalisation, de certification et de conseil en applications.

---

[107] Alexandre, L. (2018). *La guerre des intelligences: intelligence artificielle versus intelligence humaine*. Audiolib.

[108] *Handlungsspielräume in digitalen Wertschöpfungsnetzwerken* (pp. 110-119). Springer Berlin Heidelberg.

[109] Mangelsdorf, A., Wittenbrink, N., & Gabriel, P. (2022). Regulierung und Zertifizierung von KI in der Industrie: Ziele, Kriterien und Herausforderungen. In *Digitalisierung souverän gestalten II:*

Confiance.ai collabore avec VDE pour définir un label commun pour l'IA, renforçant ainsi la cohérence et la confiance dans les standards technologiques[110].

CERTAIN (DFKI) en Allemagne

CERTAIN est une initiative du DFKI axée sur la recherche, le développement, la normalisation et la certification des techniques d'IA de confiance. Ce consortium travaille sur toute la chaîne de valeur, de la recherche fondamentale à la mise en œuvre pratique, en se concentrant sur des aspects comme les explications des modèles, la causalité, la modularité, et la surveillance humaine. Confiance.ai participe activement aux projets de CERTAIN, facilitant les échanges de connaissances[111].

TAILOR en Europe

TAILOR est un projet européen visant à créer un réseau de centres d'excellence en recherche sur l'IA, combinant apprentissage, optimisation et raisonnement. Il cherche à établir les bases scientifiques d'une IA de confiance en Europe. TAILOR collabore avec Confiance.ai pour organiser des ateliers et événements scientifiques, comme la conférence ECML/PKDD[112].

RAI UK au Royaume-Uni

RAI UK est un réseau multidisciplinaire qui vise à façonner le développement de l'IA pour le bénéfice des personnes, des communautés et de la société. Il se concentre sur la création d'écosystèmes, la recherche et l'innovation, les compétences, et l'engagement du public et des politiques. Bien que Confiance.ai n'ait pas encore de collaboration établie avec RAI UK, des représentants des deux programmes participent à des événements communs[113].

CSIRO's Data61 en Australie

[110] Putzer, H. J., Rueß, H., & Koch, J. (2021). Trustworthy AI-based Systems with VDE-AR-E 2842-61. *Embedded World*.

[111] Srivastava, A., Rehm, G., & Schneider, J. M. (2017, August). DFKI-DKT at SemEval-2017 Task 8: Rumour detection and classification using cascading heuristics. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)* (pp. 486-490).

[112] Ceci, M., Hollmén, J., Todorovski, L., Vens, C., & Džeroski, S. (Eds.). (2017). *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18–22, 2017, Proceedings, Part II* (Vol. 10535). Springer.

[113] Mitchell, D., Blanche, J., Harper, S., Lim, T., Gupta, R., Zaki, O., ... & Flynn, D. (2022). A review: Challenges and opportunities for artificial intelligence and robotics in the offshore wind sector. *Energy and AI*, *8*, 100146.

Cette initiative australienne se concentre sur l'opérationnalisation de l'IA responsable en développant des technologies et outils d'ingénierie logicielle pour garantir des systèmes d'IA fiables. Data61 adopte une approche basée sur les risques et collabore étroitement avec le National AI Centre d'Australie. Confiance.ai entretient des relations régulières avec Data61, facilitant les échanges de pratiques et d'expertise[114].

Ces initiatives internationales montrent la richesse et la diversité des efforts pour développer une IA de confiance. Confiance.ai joue un rôle central en collaborant activement avec ces partenaires, en partageant des méthodologies, des technologies et des outils, et en participant à des événements internationaux. Cette coopération globale est essentielle pour garantir que les systèmes d'IA soient développés de manière éthique, fiable et bénéfique pour la société dans son ensemble.

Le programme français Confiance.AI peut tirer plusieurs leçons précieuses des initiatives internationales avec lesquelles il collabore. Voici un développement détaillé des principales leçons :

Importance de la Collaboration Internationale

La collaboration avec des programmes comme Confiance.ia au Canada, ZERTIFIZIERTE KI et CERTAIN en Allemagne, TAILOR en Europe, et CSIRO's Data61 en Australie montre que la coopération internationale est cruciale pour le développement d'une IA de confiance. Ces partenariats permettent de partager des connaissances, des méthodologies et des outils, enrichissant ainsi les capacités de Confiance.AI. Par exemple, l'interaction régulière avec des programmes canadiens et allemands a permis à Confiance.AI de bénéficier des avancées en matière de labélisation et de standardisation de l'IA. La participation à des événements communs comme le Confiance.ai Day et le symposium AITA montre l'importance de ces échanges pour harmoniser les normes et les pratiques à une échelle internationale, créant ainsi un écosystème global cohérent et fiable.

Développement de Standards et de Normes

Des initiatives comme ZERTIFIZIERTE KI mettent en avant le développement de critères d'évaluation et de standards pour l'IA. Confiance.AI peut tirer des enseignements de ces efforts en renforçant ses propres processus de standardisation. L'objectif est de garantir que les technologies d'IA soient évaluées et certifiées selon des critères stricts de fiabilité technique, d'éthique et de sécurité. Par exemple, l'expérience de ZERTIFIZIERTE KI en matière de création de mesures de sauvegarde pour atténuer les risques liés à l'IA et de développement de standards d'IA peut guider Confiance.AI dans l'élaboration de ses propres normes. En participant activement à l'élaboration de normes internationales, Confiance.AI peut s'assurer que ses critères de certification sont reconnus et adoptés globalement, renforçant ainsi la confiance dans les systèmes d'IA[115].

[114] Vitanage, D., Doolan, C., Maunsell, L., Cameron, B., Wang, Y., & Li, Z. (2018). SUCCESS IN DATA ANALYTICS-SYDNEY WATER AND DATA61 COLLABORATION. *Water e-Journal*.
[115] Gornet, M., & Maxwell, W. (2023, July). Normes techniques et éthique de l'IA. In *Conférence Nationale en Intelligence Artificielle (CNIA)*.

Approche Holistique et Interdisciplinaire

Les projets comme CERTAIN du DFKI et RAI UK soulignent l'importance d'une approche interdisciplinaire, intégrant des aspects techniques, éthiques, juridiques et sociétaux. Confiance.AI peut s'inspirer de ces initiatives pour adopter une approche plus holistique. Par exemple, CERTAIN se concentre sur des aspects clés tels que les explications des modèles, la causalité, la modularité et la surveillance humaine. Confiance.AI peut intégrer davantage d'experts en éthique, de représentants de la société civile, et de juristes dans ses comités d'évaluation pour s'assurer que tous les aspects de la confiance en l'IA sont couverts. En incorporant des perspectives variées, Confiance.AI peut mieux adresser les préoccupations des différents stakeholders et assurer que les solutions d'IA sont à la fois techniquement robustes et socialement acceptables[116].

Intégration de Critères d'Impact Social

Confiance.AI peut également apprendre de l'initiative CSIRO's Data61 en Australie, qui adopte une approche basée sur les risques tout au long du cycle de vie du développement de l'IA. L'intégration de critères d'impact social dans l'évaluation des projets d'IA est essentielle pour garantir que les technologies développées bénéficient à toutes les couches de la société. Confiance.AI peut développer des indicateurs spécifiques pour mesurer l'équité, l'inclusivité et la durabilité des technologies d'IA, s'assurant ainsi que ces technologies ne reproduisent pas ou n'amplifient pas les biais existants[117].

Formation Continue et Sensibilisation

La formation continue des professionnels de l'IA sur les aspects éthiques et légaux de leur travail est une leçon cruciale. Des initiatives comme RAI UK et CERTAIN montrent l'importance de traduire la recherche en cadres de compétences et en formations adaptées aux utilisateurs, développeurs et décideurs. Confiance.AI peut mettre en place des programmes de certification en éthique de l'IA et organiser des ateliers et des cours spécialisés pour garantir que les professionnels restent informés des meilleures pratiques et des régulations en vigueur[118].

Transparence et Responsabilité

Les audits réguliers et la transparence des processus sont également des pratiques importantes à adopter. Des audits indépendants réguliers, comme pratiqués dans les initiatives ZERTIFIZIERTE KI et CERTAIN, peuvent garantir que les systèmes d'IA respectent les normes éthiques et légales. Confiance.AI peut adopter des pratiques similaires, en publiant les

[116] Giugnatico, I. (2020). Construction d'une épistémologie critique pour la recherche interdisciplinaire à des fins d'action (intelligence artificielle et médecine personnalisée).

[117] Soriano, J. (2018). *Les enjeux de l'intégration de solutions d'intelligence artificielle au sein d'OBNL* (Doctoral dissertation, PhD thesis).

[118] COHEN, F., & ORSATELLI, P. Pour une intelligence artificielle inclusive.

résultats des audits et en assurant la transparence de ses processus d'évaluation et de certification[119].

En conclusion, le programme Confiance.AI peut tirer des leçons précieuses des initiatives internationales pour renforcer son propre cadre de développement d'une IA de confiance. La collaboration internationale, le développement de standards et de normes, une approche holistique et interdisciplinaire, l'intégration de critères d'impact social, la formation continue et la transparence sont des éléments clés à adopter. Ces leçons permettront à Confiance.AI de s'assurer que les systèmes d'IA développés en France sont fiables, éthiques et bénéfiques pour toute la société.

4.Recommandations pour les politiques et les pratiques futures

Pour renforcer et pérenniser les bénéfices du programme Confiance.AI, plusieurs recommandations détaillées peuvent être formulées afin d'orienter les politiques et les pratiques futures.

4.1. Les partenariats public-privé

Pour garantir le succès à long terme du programme Confiance.AI, il est impératif d'assurer un financement durable et de renforcer les partenariats public-privé (PPP). Le financement soutenu peut être assuré par l'allocation de fonds spécifiques à long terme par le gouvernement, combiné à des incitations fiscales pour encourager les investissements privés dans l'IA. Par exemple, des dispositifs comme le Crédit d'Impôt Recherche (CIR) peuvent être étendus pour inclure spécifiquement les projets d'IA, offrant des réductions fiscales attractives aux entreprises investissant dans la recherche et le développement de technologies d'IA[120].

En outre, la création de fonds de soutien similaires à ceux utilisés pour le développement des technologies vertes pourrait garantir un flux constant de ressources pour les projets d'IA. Ces fonds pourraient être utilisés pour financer des initiatives de recherche fondamentale ainsi que des projets applicatifs, couvrant un large éventail de secteurs, y compris la santé, les transports, et l'énergie. Les PPP jouent également un rôle crucial dans ce contexte. Ils permettent de partager les risques financiers et de tirer parti des expertises complémentaires des secteurs public et privé. Par exemple, des initiatives comme les pôles de compétitivité, qui rassemblent des entreprises, des laboratoires de recherche et des institutions académiques, peuvent stimuler l'innovation et accélérer le développement de nouvelles technologies d'IA[121].

[119] Wright, R. (2021). Sommaire Réglementer l'intelligence artificielle-Enjeux et choix essentiels.

[120] Damiano, J. P. (2020). Biomimétisme, intelligence artificielle, robotique et applications de l'intelligence en essaim. Cybersécurité et questions d'éthique et de droit. *IESF Côte d'Azur, Société des Ingénieurs et Scintifiques de France*, 7-25.

[121] SEDJARI, A. IMPACT DU NUMÉRIQUE ET DE L'INTELLIGENCE ARTIFICIELLE SUR LES TRANSFORMATIONS DES GOUVERNANCES PUBLIQUES.

Les incubateurs et les accélérateurs de start-ups spécialisés en IA représentent un autre mécanisme efficace pour soutenir l'innovation. Ces structures, soutenues par des subventions publiques et des investissements privés, offrent aux start-ups l'accès à des ressources essentielles, y compris des financements, des formations, et des réseaux de mentors et de partenaires industriels. Le succès de ces incubateurs et accélérateurs peut être vu dans des exemples comme Station F à Paris, qui héberge un écosystème vibrant de start-ups technologiques et bénéficie du soutien de grands groupes comme Facebook et Microsoft[122].

Enfin, la collaboration entre le secteur public et le secteur privé peut également inclure des initiatives de co-financement de projets de recherche en IA, où les coûts et les bénéfices sont partagés entre les partenaires. Des appels à projets conjoints et des programmes de financement collaboratif peuvent encourager l'innovation tout en répartissant les risques financiers. En conclusion, un financement durable et des partenariats public-privé solides sont essentiels pour maximiser le potentiel du programme Confiance.AI et garantir le développement continu d'une IA éthique et innovante en France.

4.2. Une régulation flexible et adaptative

Pour que le programme Confiance.AI reste pertinent et efficace, il est crucial de développer une régulation flexible et adaptative capable de suivre le rythme rapide des avancées technologiques en matière d'intelligence artificielle (IA). Ainsi, Wiener et sa cybernétique offre des pistes de réflexion intéressantes. La régulation doit être suffisamment souple pour évoluer avec les innovations tout en garantissant la sécurité, l'éthique et la protection des utilisateurs.

Tout d'abord, les régulateurs doivent élaborer des cadres législatifs qui peuvent être rapidement ajustés en réponse aux nouvelles technologies et aux évolutions du marché. Une approche "régulation en bac à sable" peut être adoptée, où les nouvelles technologies d'IA sont testées dans des environnements contrôlés avant d'être largement déployées. Cela permet aux régulateurs de comprendre les implications de ces technologies et de développer des réglementations appropriées sans freiner l'innovation. Un exemple de cette approche est la création de bacs à sable réglementaires pour les fintechs, qui a permis de tester et de réguler de nouvelles solutions financières de manière agile et efficace[123].

De plus, pour assurer la sécurité et l'éthique des applications de l'IA, il est essentiel de mettre en place des protocoles de test rigoureux. Ces protocoles doivent inclure des évaluations de l'impact éthique et des tests de robustesse pour garantir que les systèmes d'IA fonctionnent correctement et de manière sécurisée dans diverses conditions. Les tests doivent également vérifier que les algorithmes respectent les principes de non-discrimination et d'équité. La mise

[122] Soulez, M., à la Cour d'appel, A., & de Paris, L. A. B. A. (2018). Questions juridiques au sujet de l'intelligence artificielle. *Enjeux numériques*, (1), 81-85.

[123] Idoux, P. (2022). LE TEMPS DE LA REGULATION. *Le temps en droit administratif*.

en place de certifications et de labels de qualité pour les technologies d'IA pourrait également aider à rassurer les utilisateurs sur la fiabilité et la sécurité des solutions proposées[124].

La définition de normes de sécurité et de transparence est un autre aspect clé d'une régulation flexible et adaptative. Les normes de sécurité doivent garantir que les systèmes d'IA sont protégés contre les cyberattaques et les manipulations malveillantes. Par ailleurs, la transparence des algorithmes est cruciale pour garantir que les décisions prises par les systèmes d'IA peuvent être expliquées et comprises par les utilisateurs. La transparence algorithmique peut inclure des exigences de documentation détaillée des modèles d'IA, des audits indépendants réguliers et la divulgation des sources de données utilisées pour entraîner les algorithmes[125].

La Commission nationale de l'informatique et des libertés (CNIL) peut jouer un rôle central dans l'élaboration et la supervision des normes de régulation. La CNIL est déjà bien positionnée pour protéger les données personnelles et garantir la conformité aux réglementations sur la vie privée. En élargissant son rôle pour inclure la régulation des technologies d'IA, la CNIL peut s'assurer que les systèmes d'IA déployés respectent les règles de protection des données et les principes éthiques. Cela inclut la surveillance des pratiques de traitement des données, l'évaluation des impacts sur la vie privée et la mise en place de mécanismes de recours pour les individus affectés par des décisions automatisées[126].

Enfin, une régulation flexible et adaptative doit être proactive plutôt que réactive. Les régulateurs doivent anticiper les tendances technologiques et les défis potentiels pour adapter les cadres réglementaires en conséquence. Cela peut inclure la mise en place de comités d'experts multidisciplinaires pour surveiller les développements technologiques et conseiller les régulateurs sur les ajustements nécessaires. Par exemple, des groupes de travail réunissant des experts en IA, des juristes, des éthiciens et des représentants de la société civile pourraient être formés pour évaluer régulièrement l'impact des nouvelles technologies et proposer des adaptations réglementaires[127].

En conclusion, une régulation flexible et adaptative est essentielle pour accompagner le développement rapide de l'IA tout en garantissant la sécurité, l'éthique et la protection des utilisateurs. Des cadres législatifs souples, des protocoles de test rigoureux, des normes de sécurité et de transparence, le rôle central de la CNIL et une adaptation proactive sont des éléments clés pour réussir cette régulation.

---

[124] Marty, F. (2017). Algorithmes de prix, intelligence artificielle et équilibres collusifs 1. *Revue internationale de droit économique*, *31*(2), 83-116.

[125] Goffi, E. (2022). L'institutionnalisation des normes éthiques dans le domaine de l'intelligence artificielle: une construction sociale à dépasser. *Revue Ethique et Numérique*, *1*(1), 18-35.

[126] Donnat, F. (2019). L'intelligence artificielle, un danger pour la vie privée?. *Pouvoirs*, (3), 95-103.

[127] Mercier, C. (2023, October). Optimisation pédagogique grâce à l'Intelligence Artificielle: Un avenir éducatif innovant. In *Journée Pédagogique Régionale*.

4.3. Renforcement des mécanismes de gouvernance éthique

Pour garantir l'éthique et la responsabilité dans le développement et l'utilisation de l'intelligence artificielle (IA), il est crucial de renforcer les mécanismes de gouvernance éthique. La mise en place de comités d'éthique indépendants pour évaluer les projets d'IA constitue une première étape essentielle. Ces comités doivent être composés d'experts en éthique, de représentants de la société civile, de scientifiques et de juristes pour garantir une évaluation équilibrée et complète des risques et des bénéfices associés à l'IA. Ces comités peuvent fournir des recommandations et des lignes directrices pour s'assurer que les projets d'IA respectent les valeurs éthiques fondamentales et protègent les droits de l'homme.

En outre, il est important d'établir des lignes directrices claires pour l'utilisation responsable de l'IA. Ces directives devraient couvrir des aspects tels que la transparence, la responsabilité, l'équité et la protection des données. Par exemple, des normes précises pour la transparence algorithmique peuvent être établies, exigeant que les décisions prises par les systèmes d'IA soient explicables et compréhensibles par les utilisateurs. Cela inclut la documentation détaillée des modèles d'IA, des audits indépendants réguliers et la divulgation des sources de données utilisées pour entraîner les algorithmes[128].

La mise en place de systèmes de surveillance pour détecter et prévenir les abus est également cruciale. Cela peut inclure des mécanismes de contrôle continu et des audits réguliers pour s'assurer que les systèmes d'IA fonctionnent conformément aux normes éthiques et légales. Par exemple, des audits d'impact éthique et des évaluations de conformité pourraient être effectués périodiquement pour vérifier que les projets respectent les principes éthiques établis. Ces systèmes de surveillance doivent être transparents et inclure des mécanismes de recours pour les individus affectés par des décisions automatisées[129].

La transparence dans les décisions prises par les comités d'éthique et la communication régulière avec le public sont également essentielles pour maintenir la confiance. Les décisions et les recommandations des comités d'éthique devraient être publiées et accessibles au public. Des forums de discussion et des consultations publiques peuvent être organisés pour recueillir les avis et les préoccupations des citoyens sur les projets d'IA, assurant ainsi une plus grande transparence et une meilleure acceptation des technologies par le public[130].

Enfin, il est crucial d'inclure des programmes de formation continue pour les professionnels de l'IA sur les aspects éthiques et légaux de leur travail. Cela garantit que les développeurs, les ingénieurs et les gestionnaires de projets d'IA restent informés des meilleures pratiques et des régulations en vigueur. Des certifications en éthique de l'IA peuvent être proposées pour

[128] Voarino, N. (2020). Systèmes d'intelligence artificielle et santé: les enjeux d'une innovation responsable.

[129] Béranger, J., & Chassang, G. (2021). Vers une régulation néodarwinienne de l'Intelligence artificielle centrée sur le concept de l'«Ethics by Evolution». *Éthique publique. Revue internationale d'éthique sociétale et gouvernementale*, *23*(2).

[130] Meneceur, Y. (2020). La critique de la technique: clé du développement de l'intelligence artificielle?. *Les Temps Électriques en septembre*, *2020*, 42.

s'assurer que les professionnels possèdent les compétences nécessaires pour développer et déployer des systèmes d'IA de manière éthique et responsable[131].

4.4. Approche inclusive et participative

Pour garantir que les technologies d'intelligence artificielle (IA) développées et déployées répondent aux besoins et aux préoccupations de la société, il est essentiel d'adopter une approche inclusive et participative. Cette approche implique l'engagement de toutes les parties prenantes dans le processus décisionnel, y compris les citoyens, les organisations de la société civile, les experts en éthique, les chercheurs, les entreprises et les régulateurs. Une des premières étapes pour favoriser cette inclusion est d'organiser des consultations publiques régulières où les citoyens peuvent exprimer leurs opinions et préoccupations concernant les projets d'IA en cours. Par exemple, des forums de discussion, des ateliers participatifs et des plateformes en ligne peuvent être mis en place pour recueillir les avis du public. Ces initiatives permettent de s'assurer que les perspectives et les besoins de diverses communautés sont pris en compte dans le développement des technologies d'IA[132].

En parallèle, la participation active des organisations de la société civile, telles que la Ligue des droits de l'homme (LDH) et Amnesty International, est cruciale pour garantir que les projets d'IA respectent les droits de l'homme et les principes éthiques. Ces organisations peuvent fournir des perspectives essentielles sur les impacts sociaux et éthiques des technologies d'IA, contribuant à l'élaboration de politiques et de régulations qui protègent les droits individuels et collectifs. Par exemple, la LDH pourrait participer à des comités consultatifs ou des groupes de travail pour évaluer les implications des technologies d'IA sur la vie privée et la surveillance[133].

Les experts en éthique et les chercheurs doivent également être inclus dans ce processus pour fournir des analyses approfondies et des recommandations basées sur des recherches rigoureuses. Les institutions académiques et de recherche, comme l'Institut National de Recherche en Informatique et en Automatique (INRIA), peuvent jouer un rôle clé en menant des études sur les impacts sociaux et éthiques des technologies d'IA et en partageant leurs résultats avec les décideurs politiques et le public.

De plus, les entreprises développant des technologies d'IA doivent être encouragées à adopter des pratiques transparentes et responsables. Elles peuvent contribuer à cette approche inclusive en publiant des rapports de transparence, en engageant des dialogues avec les parties prenantes

[131] Leroy, C. (2023). Éthique de l'intelligence artificielle et transhumanisme: l'immortalité fait-elle la vie? 1. *Le Philosophoire*, *60*(2), 113-132.

[132] Salomé, S., & Monfort, E. (2023). Révolution numérique et âgisme: les enjeux éthiques de l'intelligence artificielle pour les personnes âgées. *NPG Neurologie-Psychiatrie-Gériatrie*, *23*(138), 383-387.

[133] RARHOUI, K. (2023). Droit de l'Intelligence Artificielle et Administration publique. *Revue Internationale du Chercheur*, *4*(4).

et en participant à des initiatives de co-régulation avec les autorités publiques. Les régulateurs, tels que la Commission nationale de l'informatique et des libertés (CNIL), peuvent faciliter cette coopération en établissant des lignes directrices claires et en organisant des tables rondes pour discuter des défis et des opportunités liés à l'IA[134].

Enfin, l'éducation et la sensibilisation du public sont des éléments clés d'une approche inclusive et participative. Des programmes éducatifs sur l'IA et ses implications éthiques doivent être développés et mis à disposition des citoyens pour les informer et les préparer à participer activement aux discussions sur l'IA. Par exemple, des cours en ligne ouverts à tous (MOOCs), des ateliers communautaires et des ressources pédagogiques accessibles peuvent aider à démystifier l'IA et à promouvoir une culture de compréhension et de responsabilité partagée[135].

4.5. Renforcement de la coopération internationale

Le renforcement de la coopération internationale est essentiel pour garantir le développement harmonisé et éthique de l'intelligence artificielle (IA) à l'échelle mondiale. Les défis posés par l'IA, tels que la protection des données, la cybersécurité, et les implications éthiques, transcendent les frontières nationales et nécessitent des réponses coordonnées. Une première étape clé est la collaboration avec des organisations internationales telles que l'Union européenne (UE), l'Organisation des Nations Unies (ONU) et l'Organisation de coopération et de développement économiques (OCDE). Par exemple, la France pourrait jouer un rôle de premier plan dans l'élaboration de cadres réglementaires européens harmonisés pour l'IA, en contribuant activement aux discussions sur la régulation de l'IA au sein de l'UE[136].

Des initiatives comme le Partenariat mondial sur l'intelligence artificielle (GPAI) permettent de rassembler des pays et des experts pour partager les meilleures pratiques et élaborer des normes communes. La France, en tant que membre fondateur, peut utiliser cette plateforme pour promouvoir des standards éthiques et techniques élevés dans le développement de l'IA, tout en bénéficiant des recherches et des innovations menées par d'autres pays . En outre, des accords bilatéraux avec des nations technologiquement avancées, comme les États-Unis, le Canada, et le Japon, peuvent faciliter l'échange de connaissances, la coopération en matière de recherche, et le développement conjoint de technologies d'IA[137].

---

[134] Nicolas, N. (2023). Kant et la morale des machines: programmer la «bonté» dans l'ère des modèles de langage. *Management & Datascience*, *7*(4).

[135] BENRAOUI, A., Khalid, E. S., & LAHMOUCHI, M. (2023). le rôle de l'innovation technologique et l'intelligence artificielle dans le renforcement de la résilience des entreprises en période de crise. *Revue Française d'Economie et de Gestion*, *4*(9).

[136] Miailhe, N. (2018). Géopolitique de l'Intelligence artificielle: le retour des empires?. *Politique étrangère*, (3), 105-117.

[137] Hadjitchoneva, J. (2020). L'intelligence artificielle au service de la prise de décisions plus efficace. *Pour une recherche économique efficace*, *149*.

La création de consortiums internationaux de recherche en IA peut également être un moyen efficace de renforcer la coopération. Ces consortiums peuvent réunir des universités, des instituts de recherche, et des entreprises de différents pays pour travailler sur des projets communs, partager des ressources et des infrastructures, et accélérer l'innovation. Par exemple, un consortium dédié à la recherche sur l'IA éthique pourrait développer des solutions technologiques tout en assurant le respect des valeurs et des normes éthiques universelles[138].

Les échanges de bonnes pratiques et de régulations à travers des conférences internationales et des workshops permettent aussi de maintenir un dialogue continu entre les différentes parties prenantes. Par exemple, des forums tels que la Conférence mondiale sur l'intelligence artificielle (World AI Conference) peuvent servir de plateformes pour discuter des défis et des opportunités liés à l'IA, promouvoir la transparence, et faciliter l'élaboration de politiques cohérentes et harmonisées[139].

Enfin, la mise en place de programmes de formation et de mobilité internationale pour les chercheurs et les professionnels de l'IA peut renforcer les compétences et encourager l'échange de savoir-faire. Des initiatives comme des programmes de bourses, des stages internationaux, et des collaborations académiques permettent de créer un réseau mondial de talents en IA, favorisant ainsi une diffusion plus rapide et plus équitable des avancées technologiques. Par exemple, des échanges entre les laboratoires de recherche français et des institutions renommées telles que le Massachusetts Institute of Technology (MIT) aux États-Unis peuvent enrichir les connaissances et stimuler l'innovation.

4.6. Mesures supplémentaires pour une IA responsable

Pour garantir le développement d'une intelligence artificielle (IA) responsable, plusieurs mesures supplémentaires doivent être mises en place, au-delà des cadres réglementaires et des initiatives de gouvernance éthique déjà mentionnés. Tout d'abord, l'intégration de critères d'impact social dans les processus d'évaluation des projets d'IA est essentielle. Cela signifie que chaque projet doit être évalué non seulement en termes de viabilité technologique et économique, mais aussi en fonction de ses impacts sociaux, économiques et environnementaux. Par exemple, des indicateurs spécifiques pourraient être développés pour mesurer l'équité, l'inclusivité et la durabilité des technologies d'IA, garantissant ainsi qu'elles bénéficient à toutes les couches de la société[140].

Des audits réguliers des systèmes d'IA pour vérifier leur conformité aux normes éthiques et légales sont également cruciaux. Ces audits doivent être menés par des tiers indépendants pour assurer l'impartialité et la rigueur des évaluations. Les résultats de ces audits devraient être

[138] Shiohira, K. (2021). Comprendre l'impact de l'intelligence artificielle sur le développement des compétences.

[139] Bruna, M. G. (2019). Quelques thèses autour de la thématique de l'Intelligence Artificielle. *Question (s) de management*, (1), 157-162.

[140] Dilhac, M. A., Abrassart, C., & Voarino, N. (2018). Rapport de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle.

rendus publics pour maintenir la transparence et la confiance du public. Par exemple, des entreprises technologiques comme Google et Microsoft pourraient être tenues de publier des rapports annuels sur l'impact éthique de leurs systèmes d'IA, détaillant les mesures prises pour éviter les biais algorithmiques et protéger les données personnelles.

Les entreprises et les organisations utilisant l'IA pourraient également être encouragées à adopter des chartes éthiques et à publier des rapports de transparence sur leurs pratiques en matière d'IA. Ces chartes devraient inclure des engagements clairs en matière de respect des droits de l'homme, de protection de la vie privée et de promotion de la diversité et de l'inclusion. Par exemple, une charte éthique pourrait stipuler que l'entreprise s'engage à ne pas utiliser l'IA pour des applications de surveillance de masse ou de reconnaissance faciale sans consentement explicite[141].

La formation continue des professionnels de l'IA sur les aspects éthiques et légaux de leur travail est une autre mesure essentielle. Cela garantit que les développeurs, les ingénieurs et les gestionnaires de projets d'IA restent informés des meilleures pratiques et des régulations en vigueur. Des programmes de certification en éthique de l'IA peuvent être proposés pour s'assurer que les professionnels possèdent les compétences nécessaires pour développer et déployer des systèmes d'IA de manière éthique et responsable. Par exemple, des universités et des institutions de formation pourraient offrir des cours spécialisés sur l'éthique de l'IA, intégrant des études de cas réels et des simulations de dilemmes éthiques.

En outre, la création de mécanismes de recours pour les individus affectés par des décisions automatisées est indispensable. Ces mécanismes doivent permettre aux personnes de contester les décisions prises par les systèmes d'IA et d'obtenir des réparations en cas de préjudice. Par exemple, des plateformes en ligne pourraient être développées pour permettre aux citoyens de signaler des abus ou des erreurs liées à l'utilisation de l'IA, et de demander des enquêtes et des corrections[142].

Enfin, il est important de promouvoir une culture de responsabilité et d'engagement éthique au sein des organisations. Cela peut être réalisé en intégrant des objectifs éthiques dans les missions et les valeurs des entreprises, et en établissant des incentives pour encourager les comportements responsables. Par exemple, des récompenses et des reconnaissances pourraient être accordées aux équipes et aux individus qui démontrent un engagement exemplaire en matière d'éthique et de responsabilité sociale dans leurs projets d'IA.

En conclusion, pour assurer le développement éthique et durable de l'intelligence artificielle (IA), plusieurs recommandations clés pour les politiques et les pratiques futures doivent être mises en œuvre. Tout d'abord, un financement durable et des partenariats public-privé solides sont essentiels pour stimuler l'innovation tout en partageant les risques financiers. Deuxièmement, il est crucial de développer une régulation flexible et adaptative capable de

[141] Béasse, M. (2021). Intelligence artificielle et mesures d'audience du journalisme en ligne: pour quelle intelligence des publics?. *Les cahiers du journalisme-Recherches*, *2*(7), R25-R38.

[142] Brahmi, H., Belouali, S., Demazeau, Y., Bouchentouf, T., & Alaoui, N. H. (2023). Vers un référentiel universel pour un usage éthique de l'intelligence artificielle. *East African Journal of Information Technology*, *6*(1), 91-106.

suivre le rythme rapide des avancées technologiques, incluant des protocoles de test rigoureux et des normes de transparence. Troisièmement, le renforcement des mécanismes de gouvernance éthique, via des comités d'éthique indépendants et des audits réguliers, est indispensable pour garantir que les projets d'IA respectent les valeurs fondamentales et les droits de l'homme. En outre, une approche inclusive et participative doit être adoptée, impliquant toutes les parties prenantes, y compris les citoyens, pour assurer que l'IA répond aux besoins et aux préoccupations de la société. Enfin, la coopération internationale doit être renforcée pour harmoniser les normes et partager les meilleures pratiques, et des mesures supplémentaires doivent être prises pour garantir une IA responsable, telles que l'intégration de critères d'impact social et la formation continue des professionnels. Ensemble, ces recommandations visent à maximiser les bénéfices de l'IA tout en minimisant les risques, assurant ainsi un avenir technologique éthique et équitable.

IV. Contributions Théoriques et Pratiques

1.Apports à la théorie de la confiance en technologie

La théorie de la confiance en technologie explore les facteurs et mécanismes qui conduisent les utilisateurs à accepter et utiliser les technologies de manière confiante et sans crainte. Dans le contexte de l'IA et des technologies de défense, cette théorie revêt une importance cruciale, car elle touche à des domaines où les conséquences d'une défaillance technologique peuvent être graves.

1.1. Fiabilité et sécurité

La fiabilité et la sécurité des systèmes technologiques sont des piliers essentiels pour instaurer la confiance, particulièrement dans le secteur de la défense où les implications de toute défaillance peuvent être critiques. La fiabilité se réfère à la capacité d'un système à fonctionner correctement et sans interruption pendant une période donnée et dans des conditions spécifiques. Cela inclut la résistance aux pannes, la robustesse face aux variations des conditions opérationnelles, et la cohérence des performances. Dans le cadre du programme Confiance.AI en France, cette fiabilité est assurée par des processus rigoureux de validation et de vérification qui testent les systèmes d'IA dans des environnements simulés avant leur déploiement réel. Ces tests incluent des scénarios de stress et des attaques simulées pour évaluer la capacité des systèmes à maintenir leurs performances face à des défis inattendus[143].

La sécurité, en parallèle, englobe la protection des systèmes contre les menaces internes et externes, notamment les cyberattaques, le sabotage et les vulnérabilités systémiques. La sécurité des systèmes d'IA dans la défense requiert une approche multi-niveau comprenant la

[143] Krichen, M. (2023). Renforcer la sécurité des contrats intelligents grâce à la puissance de l'intelligence artificielle.

sécurisation des données, la protection des algorithmes, et la mise en place de protocoles de réponse rapide aux incidents. Le programme Confiance.AI met en œuvre des pratiques de sécurité par conception, intégrant des mesures de sécurité dès les premières phases du développement des systèmes. Ces mesures incluent l'utilisation de techniques de chiffrement avancées, des audits de sécurité réguliers, et la collaboration avec des experts en cybersécurité pour identifier et corriger les failles potentielles avant qu'elles ne puissent être exploitées[144].

En outre, l'évaluation continue des risques et la mise à jour des protocoles de sécurité sont cruciales pour s'adapter à l'évolution constante des menaces. La création de redondances et de systèmes de secours contribue également à la résilience globale, permettant aux systèmes de maintenir leurs fonctions essentielles même en cas de compromission partielle. Ainsi, la combinaison de la fiabilité technique et de la robustesse des mesures de sécurité forge une base solide pour la confiance dans les technologies de défense, garantissant qu'elles peuvent non seulement performer efficacement mais aussi protéger les intérêts stratégiques et la sécurité nationale dans des environnements de plus en plus complexes et menaçants.

1.2. Transparence et explicabilité

La transparence et l'explicabilité sont des éléments cruciaux pour établir la confiance dans les systèmes d'intelligence artificielle (IA), en particulier dans le domaine de la défense. La transparence implique que les utilisateurs et les parties prenantes puissent comprendre comment les systèmes d'IA fonctionnent, quelles données ils utilisent, et comment ils prennent des décisions. Cela est essentiel pour garantir que ces systèmes soient utilisés de manière responsable et éthique. Dans le cadre du programme Confiance.AI en France, la transparence est assurée par la documentation détaillée des algorithmes et des processus décisionnels, ainsi que par la communication claire des capacités et des limites des systèmes d'IA aux utilisateurs finaux[145].

L'explicabilité, quant à elle, se réfère à la capacité d'un système d'IA à fournir des explications compréhensibles sur ses décisions et actions. Cela est particulièrement important dans les applications militaires où les décisions automatisées peuvent avoir des implications critiques. Un système d'IA explicable permet aux opérateurs humains de vérifier et de comprendre les raisons derrière une décision spécifique, ce qui est essentiel pour maintenir le contrôle humain et assurer que les décisions prises par les machines alignent avec les règles d'engagement et les principes éthiques. Confiance.AI met en œuvre des outils d'explicabilité qui utilisent des techniques comme les modèles interprétables, les visualisations de données, et les interfaces

[144] UAZRAOUI, N. (2014). *Application des Techniques de l'Intelligence Artificielle aux Problèmes de Gestion des Risques Industriels* (Doctoral dissertation, Université de Batna 2).

[145] Castets-Renard, C. (2018). Régulation des algorithmes et gouvernance du machine learning: vers une transparence et «explicabilité» des décisions algorithmiques?(Algorithm Regulation and Machine Learning Governance: Towards Transparency and'Explainability'of Algorithmic Decisions?). *Revue Droit & Affaires, Revue Paris II Assas, 15ème édition*.

utilisateur intuitives pour rendre les processus décisionnels des systèmes d'IA accessibles et compréhensibles pour les non-experts[146].

En outre, l'intégration de l'explicabilité dans les systèmes d'IA aide à détecter et à corriger les biais potentiels dans les algorithmes, renforçant ainsi la justice et l'équité dans les décisions prises. Cela permet également de renforcer la responsabilité, car les opérateurs peuvent tracer et auditer les décisions des systèmes d'IA, assurant que toute erreur ou comportement indésirable peut être rapidement identifié et rectifié. La transparence et l'explicabilité ne sont pas seulement des exigences techniques, mais elles jouent également un rôle crucial dans l'acceptation et la confiance des utilisateurs envers les technologies d'IA. En rendant les systèmes plus ouverts et compréhensibles, Confiance.AI contribue à créer un environnement où les technologies avancées peuvent être déployées en toute sécurité et avec une confiance accrue de la part des utilisateurs et des parties prenantes[147].

1.3. Ethique et régulation

Comme Wiener a pu le démontrer, l'éthique et la régulation sont des piliers essentiels pour assurer une utilisation responsable et sécurisée des systèmes d'intelligence artificielle (IA) dans le domaine de la défense. L'intégration de principes éthiques dans le développement et le déploiement des technologies d'IA est cruciale pour garantir que ces systèmes respectent les droits humains, les normes internationales et les valeurs sociétales. L'éthique dans l'IA de défense implique des considérations sur la responsabilité, la transparence, la non-discrimination, et la protection de la vie privée. Les systèmes d'IA doivent être conçus pour éviter les biais discriminatoires et garantir des décisions justes et équitables. Par exemple, le programme Confiance.AI en France inclut des comités éthiques qui supervisent le développement des technologies pour s'assurer qu'elles respectent des standards rigoureux d'éthique et de sécurité[148].

La régulation joue un rôle complémentaire en fournissant un cadre juridique et normatif qui encadre l'utilisation des technologies d'IA. Des régulations claires et bien définies aident à prévenir les abus, à protéger les droits individuels et à assurer que les technologies d'IA sont utilisées de manière responsable. En Europe, l'AI Act proposé par la Commission Européenne vise à instaurer une régulation harmonisée de l'IA, classant les systèmes selon leur niveau de risque et imposant des exigences strictes pour les applications à haut risque, notamment celles utilisées dans la défense. Cette régulation prévoit des évaluations d'impact sur la protection des

[146] Thiry, O., & Strowel, A. Le profilage à l'aide de techniques algorithmiques par les autorités publiques: la transparence et l'explicabilité, des garanties impossibles à assurer?.

[147] Delaforge, A., Azé, J., Bringay, S., Sallaberry, A., & Servajean, M. (2023, July). Panorama des outils de visualisation pour l'explicabilité en apprentissage profond pour le traitement automatique de la langue. In *CNIA 2023-Conférence Nationale en Intelligence Artificielle* (pp. 99-108).

[148] Michaud, T. Vers un filtre éthique pour les intelligences artificielles chargées de la vente?.

données et la conformité aux normes éthiques avant le déploiement de systèmes d'IA, garantissant ainsi que ces technologies sont utilisées de manière sécurisée et responsable[149].

Le programme Confiance.AI s'engage également à suivre des lignes directrices éthiques et réglementaires strictes. Il intègre des audits réguliers et des évaluations de conformité pour s'assurer que les systèmes développés respectent les régulations en vigueur. En collaborant avec des régulateurs et des experts en éthique, Confiance.AI développe des standards et des pratiques exemplaires qui peuvent être adoptés à une échelle internationale. Cela inclut la promotion de la transparence et de l'explicabilité des systèmes d'IA, ainsi que l'établissement de mécanismes de supervision et de responsabilité pour surveiller et contrôler l'utilisation de l'IA dans les opérations militaires[150].

En conclusion, l'intégration de l'éthique et de la régulation dans les systèmes d'IA de défense est indispensable pour construire et maintenir la confiance des utilisateurs et du public. Cela assure que les technologies sont développées et utilisées d'une manière qui respecte les valeurs sociétales, protège les droits humains, et garantit la sécurité nationale. Les initiatives comme Confiance.AI illustrent comment une approche rigoureuse et collaborative peut aider à naviguer les défis complexes posés par l'IA dans la défense, en équilibrant innovation technologique et responsabilité éthique.

1.4. Collaborations et standards internationaux

Les collaborations internationales et l'établissement de standards globaux jouent un rôle crucial dans le développement et l'implémentation responsables des systèmes d'intelligence artificielle (IA) dans le domaine de la défense. Ces collaborations permettent de partager les meilleures pratiques, d'harmoniser les normes et d'assurer que les technologies d'IA sont déployées de manière cohérente et sécurisée à travers le monde. Le programme Confiance.AI en France, par exemple, travaille en étroite collaboration avec des initiatives internationales comme le projet ZERTIFIZIERTE KI en Allemagne et le réseau européen TAILOR, qui rassemble des chercheurs et des industriels pour développer des standards communs en matière d'IA de confiance. Ces collaborations facilitent l'échange de connaissances et l'alignement des efforts de recherche et développement, renforçant ainsi la capacité collective à faire face aux défis éthiques, techniques et sécuritaires posés par l'IA[151].

L'établissement de standards internationaux est essentiel pour garantir que les systèmes d'IA respectent des exigences uniformes en termes de fiabilité, de sécurité et d'éthique. Les standards

[149] Goffi, E. (2022). L'institutionnalisation des normes éthiques dans le domaine de l'intelligence artificielle: une construction sociale à dépasser. *Revue Ethique et Numérique*, *1*(1), 18-35.

[150] Goffi, E. (2022). L'institutionnalisation des normes éthiques dans le domaine de l'intelligence artificielle: une construction sociale à dépasser. *Revue Ethique et Numérique*, *1*(1), 18-35.

[151] Senneville-Robert, M. (2021). *Reconfiguration des Liens de Collaboration Entre Acteurs Industriels et Universitaires de la Recherche en Intelligence Artificielle à Montréal et à Toronto* (Doctoral dissertation, Institut National de la Recherche Scientifique (Canada)).

permettent de créer un cadre de référence commun qui aide les développeurs et les utilisateurs à évaluer la conformité et la performance des technologies d'IA. Par exemple, les lignes directrices de l'Organisation internationale de normalisation (ISO) et les recommandations de l'Institut des ingénieurs électriciens et électroniciens (IEEE) fournissent des bases solides pour le développement de systèmes d'IA sécurisés et éthiques. Ces standards couvrent divers aspects, tels que la gestion des données, la robustesse des algorithmes, et la transparence des processus décisionnels[152].

La collaboration transnationale est également vitale pour aborder les questions de régulation et de gouvernance. Les régulations nationales peuvent varier considérablement, et une approche harmonisée facilite la mise en œuvre de régulations cohérentes et efficaces. Les initiatives comme l'AI Act de la Commission Européenne visent à créer un cadre réglementaire unifié qui peut servir de modèle pour d'autres régions du monde, en promouvant une IA fiable et sécurisée. En travaillant ensemble, les pays peuvent développer des régulations qui non seulement protègent les utilisateurs mais aussi favorisent l'innovation responsable et éthique.

En résumé, les collaborations internationales et l'établissement de standards globaux sont essentiels pour garantir que les systèmes d'IA de défense sont développés et déployés de manière sécurisée, éthique et cohérente. Ces efforts collectifs permettent de maximiser les bénéfices des technologies d'IA tout en minimisant les risques, en assurant que les avancées technologiques servent l'intérêt général et respectent les valeurs fondamentales des sociétés à travers le monde[153].

En résumé, les apports à la théorie de la confiance en technologie dans le contexte des systèmes d'IA et de défense incluent une attention accrue à la fiabilité, à la transparence, à l'éthique, à la régulation adaptative, et à la collaboration internationale. Ces éléments sont essentiels pour développer des technologies qui non seulement fonctionnent efficacement mais sont également dignes de confiance et alignées sur les valeurs sociétales et éthiques.

2.Implications pour la pratique dans le domaine de la défense et de l'IA

L'intégration de l'intelligence artificielle (IA) dans le domaine de la défense engendre des implications profondes et variées pour les pratiques militaires et de sécurité nationale. Ces implications touchent plusieurs aspects clés, allant de l'optimisation des opérations à la transformation des stratégies de défense, en passant par l'évolution des rôles humains dans les environnements militaires.

2.1. Optimisation des opérations militaires

[152] Côté, A. M., & Su, Z. (2021). Évolutions de l'intelligence artificielle au travail et collaborations humain-machine. *Ad machina*, (5), 144-160.

[153] Romero, M., Heiser, L., & Lepage, A. (2023). Enseigner et apprendre à l'ère de l'intelligence artificielle: acculturation, intégration et usages créatifs de l'IA en éducation: livre blanc.

L'intégration de l'intelligence artificielle (IA) dans les opérations militaires offre un potentiel considérable pour optimiser les performances, la réactivité et l'efficacité des forces armées. Les systèmes d'IA peuvent traiter et analyser rapidement d'énormes volumes de données provenant de multiples sources, telles que les capteurs de surveillance, les satellites, et les rapports de renseignement. Cette capacité permet de fournir des informations critiques en temps réel, améliorant ainsi la prise de décision et la coordination des opérations. Par exemple, les algorithmes de reconnaissance d'image et de traitement du langage naturel peuvent identifier et classer des menaces potentielles, des cibles stratégiques, et des mouvements ennemis avec une précision inégalée, permettant aux commandants de réagir rapidement et de manière informée aux situations évolutives sur le terrain[154].

En outre, l'IA facilite la planification et l'exécution des missions en automatisant les tâches répétitives et en optimisant l'allocation des ressources. Les systèmes basés sur l'IA peuvent simuler différents scénarios de mission pour déterminer les stratégies les plus efficaces et prévoir les résultats possibles, réduisant ainsi les risques et augmentant les chances de succès. Par exemple, les drones autonomes peuvent être utilisés pour des missions de reconnaissance et de surveillance, fournissant des données en temps réel sans exposer les soldats à des dangers immédiats. Les véhicules autonomes peuvent également transporter des fournitures et évacuer des blessés, améliorant ainsi la logistique et la sécurité des opérations sur le terrain. Evidemment, ce genre de technologie doivent être soumises à un contrôle important. De ce fait, les Trois Lois de la Robotique d'Isaac Asimov offre un début de réflexion sur ce sujet.

Le programme Confiance.AI en France illustre bien cette optimisation des opérations militaires. En développant des systèmes d'IA robustes et fiables, le programme vise à renforcer les capacités de collecte et d'analyse de données, à améliorer la précision des frappes et des interventions, et à garantir une coordination plus fluide entre les différentes unités et branches des forces armées. En intégrant des technologies avancées telles que l'apprentissage automatique et l'intelligence artificielle dans les systèmes de commandement et de contrôle, Confiance.AI contribue à créer un environnement où les décisions peuvent être prises rapidement et efficacement, tout en minimisant les erreurs humaines et les pertes potentielles [155].

Ainsi, l'optimisation des opérations militaires grâce à l'IA permet non seulement de maximiser l'efficacité et la sécurité des missions, mais aussi de libérer le potentiel humain pour des tâches stratégiques à plus haute valeur ajoutée, créant un avantage significatif sur le champ de bataille moderne.

2.2. Transformations des stratégies de Défense

[154] Godin, V. (2021). Singularité militaire et digitalisation: les défis de l'intelligence artificielle dans l'optimisation des ressources humaines. *Revue Defense Nationale*, (HS4), 227-240.

[155] Jean-Christophe, N. O. Ë. L. (2018). Intelligence artificielle: vers une nouvelle révolution militaire?. *Focus stratégique*, (84).

L'intégration de l'intelligence artificielle (IA) transforme radicalement les stratégies de défense, redéfinissant les méthodes traditionnelles et ouvrant de nouvelles perspectives pour les opérations militaires. Les technologies d'IA permettent de développer des capacités avancées telles que les systèmes autonomes, l'analyse prédictive et la guerre cybernétique, qui renforcent la supériorité technologique et stratégique des forces armées. Par exemple, les drones autonomes et les robots peuvent être déployés pour des missions de reconnaissance, de surveillance et de logistique, réduisant les risques pour les soldats tout en augmentant la précision et l'efficacité des opérations. Ces systèmes autonomes peuvent opérer dans des environnements dangereux ou inaccessibles, fournissant des renseignements en temps réel et permettant des interventions rapides et ciblées[156].

De plus, l'IA transforme la collecte et l'analyse des renseignements militaires. Les algorithmes de machine learning peuvent traiter d'énormes volumes de données provenant de diverses sources, comme les satellites, les capteurs, et les réseaux sociaux, pour identifier des schémas et des anomalies indicatives de menaces potentielles. Cette capacité d'analyse prédictive permet aux forces armées de détecter et de neutraliser les menaces avant qu'elles ne se matérialisent, améliorant ainsi la sécurité nationale et la préparation stratégique. Le programme Confiance.AI en France illustre cette transformation en développant des outils d'IA pour améliorer la prise de décision stratégique et opérationnelle, en intégrant des technologies avancées dans les systèmes de commandement et de contrôle pour une réponse plus rapide et plus efficace aux situations de crise.

L'IA impacte également les doctrines militaires et les stratégies de combat. Les systèmes d'IA peuvent simuler différents scénarios de conflit, permettant aux planificateurs militaires de tester et d'optimiser leurs stratégies avant de les déployer sur le terrain. Ces simulations peuvent intégrer des variables complexes et changeantes, fournissant des insights précieux pour la planification et l'exécution des opérations militaires. De plus, l'IA joue un rôle crucial dans la guerre cybernétique, où les systèmes automatisés peuvent détecter, analyser et contrer les cyberattaques plus rapidement et avec plus de précision que les méthodes traditionnelles.

Enfin, l'intégration de l'IA dans les stratégies de défense nécessite une réévaluation des rôles humains et de la formation militaire. Les soldats et les opérateurs doivent être formés pour interagir efficacement avec les systèmes d'IA, comprendre leurs capacités et limitations, et prendre des décisions éclairées basées sur les informations fournies par ces technologies. Cela nécessite un investissement dans la formation et l'éducation pour développer les compétences nécessaires pour opérer dans un environnement militaire de plus en plus technologiquement avancé[157].

En somme, l'IA transforme les stratégies de défense en introduisant des capacités technologiques avancées qui augmentent l'efficacité opérationnelle, améliorent la prise de décision stratégique et rehaussent la supériorité technologique sur le champ de bataille. Le

[156] Ventre, D. (2020). *Intelligence artificielle, cybersécurité et cyberdéfense* (Vol. 2). ISTE Group.

[157] Meunier, F., & Bombal, S. (2018). Prospective sur l'intelligence artificielle au profit de la cyberdéfense. *Revue Défense Nationale*, (4), 72-77.

programme Confiance.AI incarne cette transformation en développant des systèmes d'IA robustes et éthiques, capables de répondre aux défis contemporains et futurs de la défense nationale et internationale.

2.3. Evolution des rôles humains

L'intégration de l'intelligence artificielle (IA) dans le domaine de la défense entraîne une évolution significative des rôles humains, modifiant la nature des tâches et des responsabilités des personnels militaires. Les systèmes d'IA sont capables d'automatiser de nombreuses tâches répétitives et analytiques, permettant ainsi aux soldats et aux officiers de se concentrer sur des activités à plus haute valeur ajoutée, comme la prise de décision stratégique et la gestion des opérations complexes. Par exemple, les algorithmes de traitement des données peuvent analyser des flux massifs d'informations pour identifier des menaces potentielles, laissant aux analystes humains le soin d'interpréter ces résultats et de formuler des stratégies appropriées. Cette automatisation réduit la charge cognitive des personnels et améliore l'efficacité opérationnelle en accélérant le processus de prise de décision[158].

En outre, l'introduction de l'IA modifie les compétences et les formations requises pour les militaires. Les soldats doivent désormais être formés non seulement aux compétences traditionnelles de combat, mais aussi à l'utilisation et à la gestion des technologies avancées. Cela inclut la compréhension des systèmes d'IA, de leurs capacités et de leurs limitations, ainsi que la capacité à collaborer avec des machines intelligentes dans des environnements opérationnels. Le programme Confiance.AI en France met en place des initiatives de formation pour garantir que le personnel militaire puisse utiliser efficacement les nouvelles technologies, assurant ainsi une transition en douceur vers une armée augmentée par l'IA[159].

L'évolution des rôles humains dans la défense implique également une réévaluation de la responsabilité et de la supervision. Avec l'IA prenant des décisions autonomes dans certains contextes, il est crucial de maintenir un niveau de contrôle humain significatif pour s'assurer que ces décisions respectent les règles d'engagement et les principes éthiques. Les opérateurs humains doivent être capables de comprendre et de superviser les actions des systèmes d'IA, intervenant lorsque cela est nécessaire pour corriger ou ajuster les décisions automatisées. Cette supervision est essentielle pour maintenir la confiance dans les systèmes d'IA et pour garantir que leur utilisation reste alignée avec les valeurs et les objectifs stratégiques des forces armées[160].

---

[158] Zouinar, M. (2020). Évolutions de l'Intelligence Artificielle: quels enjeux pour l'activité humaine et la relation Humain-Machine au travail?. *Activités*, (17-1).

[159] Ganascia, J. G. (2017). *Le Mythe de la Singularité. Faut-il craindre l'intelligence artificielle?*. Média Diffusion.

[160] Alexandre, L. (2018). *La guerre des intelligences: intelligence artificielle versus intelligence humaine*. Audiolib.

Enfin, l'IA offre également de nouvelles opportunités pour les rôles humains en termes de maintenance et de développement technologique. Les techniciens et les ingénieurs militaires doivent être formés pour entretenir et améliorer les systèmes d'IA, assurant leur fiabilité et leur performance optimale. Cela crée de nouveaux postes et des spécialisations au sein des forces armées, contribuant à la professionnalisation et à l'innovation continue dans le secteur de la défense.

En somme, l'évolution des rôles humains dans la défense, sous l'influence de l'IA, transforme la nature des tâches militaires, enrichit les compétences requises et redéfinit les responsabilités et la supervision. Le programme Confiance.AI illustre comment une approche intégrée de la formation et de la collaboration humain-machine peut maximiser les avantages de l'IA tout en préservant la valeur stratégique et éthique des forces armées.

2.4. Considérations éthiques et réglementaires

En somme, l'intégration de l'IA dans la défense offre des opportunités considérables pour améliorer les opérations militaires, transformer les stratégies de défense, et réorienter les rôles humains vers des tâches à plus haute valeur ajoutée. Toutefois, cela nécessite une attention continue aux considérations éthiques, une régulation appropriée, et une collaboration internationale pour assurer que les technologies d'IA sont utilisées de manière sécurisée, responsable et alignée sur les valeurs sociétales. Le programme Confiance.AI illustre comment une approche intégrée et multidisciplinaire peut aider à naviguer ces implications complexes, maximisant les bénéfices tout en minimisant les risques[161].

3.Potentiel pour des études et recherches futures

Le domaine de l'intelligence artificielle (IA) appliqué à la défense est riche en opportunités pour des études et des recherches futures. Ces recherches peuvent approfondir notre compréhension des technologies actuelles et ouvrir de nouvelles voies pour leur application, en maximisant les avantages tout en minimisant les risques associés. Voici quelques directions prometteuses pour les études et les recherches futures :

L'interopérabilité des systèmes d'IA entre différentes branches militaires et entre forces armées alliées est une priorité. La recherche future peut se concentrer sur le développement de standards communs et de protocoles de communication permettant une intégration fluide des technologies d'IA dans des opérations conjointes. Cela inclut l'élaboration de frameworks techniques et organisationnels qui facilitent l'interopérabilité, garantissant que les systèmes

[161] Azaroual, F. (2024). L'Intelligence Artificielle en Afrique: défis et opportunités.

d'IA peuvent échanger des informations et fonctionner ensemble de manière cohérente et efficace[162].

Une autre direction cruciale est l'étude de l'impact des systèmes d'IA autonomes sur la dynamique du commandement et du contrôle militaire. Les chercheurs peuvent explorer comment l'IA peut aider à automatiser et optimiser les décisions tactiques et stratégiques, tout en maintenant la supervision humaine. Cela implique la conception de systèmes d'IA explicables et transparents, qui fournissent des justifications compréhensibles pour leurs décisions, permettant ainsi aux commandants de valider et de superviser les actions automatisées. La recherche peut également aborder les mécanismes de responsabilité et de gouvernance pour garantir que les décisions prises par des systèmes autonomes respectent les normes éthiques et légales[163].

Les biais algorithmiques sont une préoccupation majeure dans le développement des systèmes d'IA. La recherche future peut se concentrer sur des méthodes pour détecter, atténuer et éliminer les biais dans les algorithmes d'IA, garantissant ainsi des décisions justes et équitables. Des études peuvent également explorer les cadres éthiques pour l'utilisation de l'IA dans la défense, en s'assurant que les systèmes respectent les droits humains et les normes internationales de droit humanitaire. L'élaboration et l'évaluation de ces cadres dans des scénarios réels permettront de développer des lignes directrices robustes pour l'utilisation éthique de l'IA.

La cybersécurité militaire est un domaine où l'IA peut avoir un impact significatif. Les systèmes d'IA peuvent détecter, analyser et contrer les cybermenaces de manière plus rapide et précise que les méthodes traditionnelles. La recherche future peut se pencher sur le développement de techniques d'apprentissage automatique capables de reconnaître des attaques nouvelles et inconnues, ainsi que sur la création de systèmes résilients capables de continuer à fonctionner même lorsqu'ils sont partiellement compromis. En outre, l'intégration de l'IA dans les infrastructures de cybersécurité existantes et l'évaluation de son efficacité constituent des domaines de recherche importants[164].

L'impact de l'IA sur le personnel militaire, tant psychologique que sociologique, est un autre domaine clé de recherche. Comprendre comment les soldats interagissent avec les systèmes d'IA, et comment ces technologies influencent leur moral, leur efficacité et leur bien-être, est crucial pour une intégration réussie. Des études peuvent être menées pour développer des programmes de formation et de soutien adaptés, visant à maximiser les bénéfices de l'IA tout en minimisant les impacts négatifs sur les individus. Cela inclut des recherches sur la

[162] Agard, G., Bianchi, A., Bernat, M., Duclos, G., & Leone, M. (2024). A tale of ChatGPT: revue de la littérature sur la place des bêtabloquants dans le choc septique par intelligence artificielle générative (ChatGPT). *Anesthésie & Réanimation*.

[163] HAMDANI, F., MONTICOLO, D., & BOLY, V. Etude de l'apport de l'Intelligence Artificielle pour l'innovation de produit.

[164] Fuhrer 1, C. (2023). Intelligence Artificielle: que dit la recherche récente? Une approche combinée bibliométrique et textuelle. *Management & Avenir*, (5), 89-111.

perception de l'IA par le personnel, leur niveau de confiance dans ces systèmes, et les stratégies pour atténuer toute résistance au changement technologique[165].

L'IA ouvre également la porte à de nouvelles applications innovantes dans le domaine militaire. La recherche peut explorer des domaines tels que l'optimisation logistique, la maintenance prédictive des équipements, et l'amélioration des capacités de formation par le biais de simulations réalistes. Les systèmes d'IA peuvent créer des environnements virtuels pour l'entraînement des soldats, permettant des simulations de combat complexes et variées qui préparent mieux les forces armées aux situations réelles. Ces technologies peuvent également aider à prévoir les besoins logistiques et à planifier les opérations de maintenance, réduisant les coûts et augmentant l'efficacité opérationnelle.

Enfin, l'intégration de l'IA dans les politiques de défense et les doctrines militaires est un domaine d'étude essentiel. Les chercheurs peuvent analyser comment les technologies d'IA influencent les stratégies de défense nationale et internationale, et comment elles peuvent être intégrées dans les politiques de sécurité pour maximiser leur impact positif. Cela inclut l'étude des implications géopolitiques de l'adoption de l'IA par différentes nations, ainsi que les stratégies de collaboration et de régulation internationales pour assurer une utilisation pacifique et sécurisée de l'IA dans la défense[166].

En conclusion, le potentiel pour des études et des recherches futures dans le domaine de l'IA de défense est vaste et multifacette. En explorant ces différentes directions, la recherche peut non seulement améliorer les capacités technologiques, mais aussi garantir une utilisation éthique et responsable de l'IA, alignée sur les valeurs et les normes internationales. Les initiatives comme le programme Confiance.AI démontrent comment une approche intégrée et multidisciplinaire peut aider à naviguer les défis complexes posés par l'IA dans la défense, maximisant ainsi les bénéfices tout en minimisant les risques.

Conclusion

Ce mémoire de recherche s'est attaché à examiner la sécurité à l'ère de l'intelligence artificielle (IA), en se concentrant sur l'établissement d'un cadre de confiance dans le secteur de la défense à travers l'analyse du programme français Confiance.AI. Les principaux résultats de cette étude montrent une approche intégrée et multidimensionnelle pour aborder les défis de l'IA dans la défense, en combinant des considérations techniques, éthiques, réglementaires et collaboratives.

Premièrement, en ce qui concerne la fiabilité et la sécurité, le programme Confiance.AI a mis en place des processus rigoureux de validation et de vérification pour assurer que les systèmes d'IA déployés respectent des normes strictes de performance et de sécurité. Ces processus

[165] Conti, S., Baudier, P., & Billot, R. (2023). Impact de l'intelligence Artificielle dans les services clients. *Management Avenir*, *137*(5), 69-88.

[166] Gauthy Choc, L. (2023). État des lieux et analyse de l'apport potentiel du Cloud Computing et des techniques de Re connaissance Optique de Caractères dans le métier d'auditeur.

incluent des tests dans des environnements simulés pour garantir que les technologies peuvent fonctionner de manière fiable et prévisible dans des conditions opérationnelles réelles. La recherche a montré que ces mesures sont essentielles pour éviter les défaillances potentielles et assurer une performance constante des systèmes d'IA, réduisant ainsi les risques associés à leur utilisation dans des contextes militaires critiques.

Deuxièmement, l'étude a mis en lumière l'importance de la transparence et de l'explicabilité des systèmes d'IA. La transparence signifie que les utilisateurs doivent être en mesure de comprendre comment et pourquoi les technologies prennent certaines décisions. Confiance.AI a développé des outils pour l'explicabilité, permettant aux opérateurs humains de vérifier et de comprendre les décisions prises par les systèmes d'IA. Cette démarche est cruciale pour maintenir la confiance des utilisateurs et garantir que les systèmes sont utilisés de manière responsable et éthique. Les algorithmes explicables aident également à détecter et à corriger les biais potentiels, renforçant ainsi la justice et l'équité des décisions prises par les machines.

Troisièmement, l'intégration de l'éthique et de la régulation dans le développement et le déploiement des technologies d'IA constitue un autre résultat majeur de cette recherche. Le programme Confiance.AI travaille avec des comités d'éthique et adopte des régulations strictes pour s'assurer que les technologies développées sont conformes aux principes éthiques et respectent les droits humains. Cette approche est renforcée par des régulations adaptatives qui évoluent avec les avancées technologiques, garantissant une utilisation responsable des systèmes d'IA. Les efforts pour intégrer des considérations éthiques dès le début du processus de développement sont essentiels pour prévenir les abus et protéger les droits individuels.

Enfin, la collaboration internationale et l'établissement de standards globaux apparaissent comme des éléments essentiels pour harmoniser les pratiques et renforcer la confiance dans les technologies d'IA à l'échelle mondiale. Confiance.AI collabore avec des initiatives internationales, telles que ZERTIFIZIERTE KI en Allemagne et TAILOR en Europe, pour partager les meilleures pratiques et développer des normes communes. Cette coopération transnationale est vitale pour s'assurer que les systèmes d'IA développés sont reconnus et acceptés internationalement, ce qui est crucial pour leur déploiement efficace et sécurisé.

La problématique centrale de ce mémoire, "Comment instaurer un cadre de confiance dans le secteur de la défense à l'ère de l'intelligence artificielle ?", a été abordée à travers l'analyse détaillée du programme Confiance.AI. Les stratégies et les pratiques adoptées par ce programme fournissent des réponses concrètes à cette question en démontrant qu'un cadre de confiance peut être instauré en combinant rigueur technique, transparence, éthique et collaboration internationale.

Les efforts de Confiance.AI pour développer des technologies fiables et explicables montrent que la confiance peut être établie en s'assurant que les systèmes d'IA fonctionnent de manière prévisible et compréhensible. En intégrant des principes éthiques dès le début du développement et en adoptant des régulations strictes, le programme garantit que les technologies sont utilisées de manière responsable et alignée sur les valeurs sociétales. De plus, la coopération internationale renforce la crédibilité et l'acceptation des technologies d'IA, facilitant leur déploiement à l'échelle mondiale.

Les implications de cette recherche soulignent la nécessité d'une approche holistique et multidimensionnelle pour gérer les défis posés par l'intégration de l'IA dans la défense. La collaboration entre les différents acteurs — gouvernements, industries, institutions de recherche et société civile — est essentielle non seulement pour le développement de technologies robustes et sécurisées, mais aussi pour l'établissement de normes éthiques et réglementaires globales qui assurent une utilisation responsable de l'IA.

Les perspectives futures pour la recherche et les pratiques dans ce domaine sont vastes. Des études supplémentaires pourraient se concentrer sur l'amélioration continue des systèmes d'IA en termes de robustesse et de sécurité. Cela inclut le développement de méthodes avancées pour l'explicabilité des décisions automatisées et l'exploration des impacts sociétaux de l'IA militaire. Par exemple, les recherches pourraient se pencher sur l'intégration de techniques de machine learning plus sophistiquées pour améliorer la détection et la prévention des cybermenaces, ainsi que sur le développement de cadres de gouvernance adaptatifs pour gérer l'évolution rapide des technologies d'IA.

L'impact psychologique et sociologique de l'IA sur le personnel militaire constitue également un domaine de recherche important. Comprendre comment les soldats interagissent avec les systèmes d'IA et comment ces technologies influencent leur moral, leur efficacité et leur bien-être est crucial pour une intégration réussie. Des études peuvent être menées pour développer des programmes de formation et de soutien adaptés, maximisant ainsi les bénéfices de l'IA tout en minimisant les impacts négatifs sur les individus.

En outre, l'optimisation des opérations militaires grâce à l'IA offre un potentiel significatif pour améliorer l'efficacité et la sécurité des missions. Les systèmes d'IA peuvent traiter et analyser rapidement de vastes quantités de données, fournissant des informations critiques en temps réel pour la prise de décision. Cela améliore la réactivité et la coordination des opérations, permettant des réponses plus rapides et mieux informées aux menaces. Les recherches futures pourraient explorer de nouvelles applications de l'IA pour l'optimisation logistique, la maintenance prédictive des équipements et l'amélioration des capacités de formation par le biais de simulations réalistes.

Enfin, l'intégration de l'IA dans les politiques de défense et les doctrines militaires est un domaine d'étude essentiel. Les chercheurs peuvent analyser comment les technologies d'IA influencent les stratégies de défense nationale et internationale, et comment elles peuvent être intégrées dans les politiques de sécurité pour maximiser leur impact positif. Cela inclut l'étude des implications géopolitiques de l'adoption de l'IA par différentes nations et les stratégies de collaboration et de régulation internationales pour assurer une utilisation pacifique et sécurisée de l'IA dans la défense.

En conclusion, le programme Confiance.AI représente une initiative exemplaire qui démontre comment un État peut aborder les défis complexes de l'IA dans la défense de manière responsable et efficace. En adoptant des pratiques innovantes et en mettant l'accent sur la transparence, l'éthique et la collaboration internationale, il est possible de développer des technologies qui non seulement renforcent les capacités de défense mais respectent également les normes éthiques et sécuritaires indispensables dans un monde de plus en plus interconnecté

et technologiquement avancé. Les résultats de cette recherche fournissent une base solide pour la réflexion et l'action future, en soulignant l'importance d'une approche intégrée et multidisciplinaire pour naviguer dans l'ère de l'intelligence artificielle avec un impact positif sur la sécurité nationale et internationale.

Références :

Agard, G., Bianchi, A., Bernat, M., Duclos, G., & Leone, M. (2024). A tale of ChatGPT: revue de la littérature sur la place des bêtabloquants dans le choc septique par intelligence artificielle générative (ChatGPT). *Anesthésie & Réanimation*.

AI for Humanity. (2020). Évaluation du programme Confiance.AI. Ministère de l'Économie et des Finances.

Alexandre, L. (2018). *La guerre des intelligences: intelligence artificielle versus intelligence humaine*. Audiolib.

Asimov, I. (1941). Three laws of robotics. *Asimov, I. Runaround*, *2*, 3.

Aubin-Auger, I., Mercier, A., Baumann, L., Lehr-Drylewicz, A. M., Imbert, P., & Letrilliart, L. (2008). Introduction à la recherche qualitative. *Exercer*, *84*(19), 142-5.

Azaroual, F. (2024). L'Intelligence Artificielle en Afrique: défis et opportunités.

Balacheff, N. (1994). Didactique et intelligence artificielle. *Recherches en didactique des mathématiques*, *14*, 9-42.

Bastian, N. D. (2020). Building the Army's Artificial Intelligence Workforce. *The Cyber Defense Review*, *5*(2), 59-64.

Béasse, M. (2021). Intelligence artificielle et mesures d'audience du journalisme en ligne: pour quelle intelligence des publics?. *Les cahiers du journalisme-Recherches*, *2*(7), R25-R38.

BENRAOUI, A., Khalid, E. S., & LAHMOUCHI, M. (2023). le rôle de l'innovation technologique et l'intelligence artificielle dans le renforcement de la résilience des entreprises en période de crise. *Revue Française d'Economie et de Gestion*, *4*(9).

Béranger, J., & Chassang, G. (2021). Vers une régulation néodarwinienne de l'Intelligence artificielle centrée sur le concept de l'«Ethics by Evolution». *Éthique publique. Revue internationale d'éthique sociétale et gouvernementale*, *23*(2).

Bezombes, P. (2018). Quelle éthique opérationnelle pour les systèmes automatisés et l'intelligence artificielle?. *Revue Défense Nationale*, (4), 89-92.

Binnendijk, A. (2020). The Role of Artificial Intelligence in Future Warfare. RAND Corporation.

Bostrom, N. (2001). What is transhumanism. *Nick Bostrom*, *1998*.

Bostrom, N. (2005). A history of transhumanist thought. *Journal of evolution and technology*, *14*(1).

Boulanin, V., & Verbruggen, M. (2017). Mapping the Development of Autonomy in Weapon Systems. SIPRI.

Bruna, M. G. (2019). Quelques thèses autour de la thématique de l'Intelligence Artificielle. *Question (s) de management*, (1), 157-162.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. arXiv preprint arXiv:1802.07228.

Brynjolfsson, E., & McAfee, A. (2014). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. W. W. Norton & Company.

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency.

Castets-Renard, C. (2018). Régulation des algorithmes et gouvernance du machine learning: vers une transparence et «explicabilité» des décisions algorithmiques?(Algorithm Regulation and Machine Learning Governance: Towards Transparency and'Explainability'of Algorithmic Decisions?). *Revue Droit & Affaires, Revue Paris II Assas, 15ème édition*.

Castillo, M. (2023). L'Union européenne: vers la maîtrise de l'intelligence artificielle?. *Cahiers de la recherche sur les droits fondamentaux*, (21), 99-107.

Ceci, M., Hollmén, J., Todorovski, L., Vens, C., & Džeroski, S. (Eds.). (2017). *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18–22, 2017, Proceedings, Part II* (Vol. 10535). Springer.

Chiva, E. (2019). L'intelligence artificielle: un moteur de l'innovation de défense française. *Revue Défense Nationale*, (5), 33-37.

COHEN, F., & ORSATELLI, P. Pour une intelligence artificielle inclusive.

Commission Européenne (2019). Ethics Guidelines for Trustworthy AI.

Commission Européenne. (2019). Lignes directrices pour une IA digne de confiance.

Commission Européenne. (2020). Lignes directrices pour une IA digne de confiance.

Confiance.ai. (2021). Programme de développement et d'intégration de l'IA.

Côté, A. M., & Su, Z. (2021). Évolutions de l'intelligence artificielle au travail et collaborations humain-machine. *Ad machina*, (5), 144-160.

Crootof, R. (2015). The Killer Robots Are Here: Legal and Policy Implications. Cardozo Law Review, 36(1), 183-201.

Cugurullo, F. (2021). «One AI to rule them all». En Chine, l'unification de la gouvernance urbaine par l'intelligence artificielle. *GREEN*, (1), 134-137.

Cummings, M. L. (2017). Artificial Intelligence and the Future of Warfare. Chatham House.

Damiano, J. P. (2020). Biomimétisme, intelligence artificielle, robotique et applications de l'intelligence en essaim. Cybersécurité et questions d'éthique et de droit. *IESF Côte d'Azur, Société des Ingénieurs et Scintifiques de France*, 7-25.

Defense Advanced Research Projects Agency. (2020). Artificial Intelligence Explorations.

Delaforge, A., Azé, J., Bringay, S., Sallaberry, A., & Servajean, M. (2023, July). Panorama des outils de visualisation pour l'explicabilité en apprentissage profond pour le traitement automatique de la langue. In *CNIA 2023-Conférence Nationale en Intelligence Artificielle* (pp. 99-108).

Deslauriers, J. P. (1987). Lfanalyse en recherche qualitative. *Cahiers de recherche sociologique*, *5*(2), 145-152.

Dilhac, M. A., Abrassart, C., & Voarino, N. (2018). Rapport de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle.

Donnat, F. (2019). L'intelligence artificielle, un danger pour la vie privée?. *Pouvoirs*, (3), 95-103.

Doshi-Velez, F., & Kim, B. (2017). Towards a Rigorous Science of Interpretable Machine Learning. arXiv preprint arXiv:1702.08608.

Dumez, H. (2011). Faire une revue de littérature: pourquoi et comment?. *Le libellio d'aegis*, *7*(2-Eté), 15-27.

Dumez, H. (2011). Qu'est-ce que la recherche qualitative?. *Le Libellio d'Aegis*, *7*(4-Hiver), 47-58.

ENISA (2021). Artificial Intelligence Cybersecurity Challenges.

European Commission. (2020). White Paper on Artificial Intelligence: A European approach to excellence and trust. European Commission.

Ezzaher, R., & Bouklata, A. (2020, December). Méthodologie d'évaluation de l'impact du transport de marchandises en ville sur le développement durable: Revue de littérature. In *2020 IEEE 13th International Colloquium of Logistics and Supply Chain Management (LOGISTIQUA)* (pp. 1-7). IEEE.

Federal Trade Commission. (2020). Using Artificial Intelligence and Algorithms.

Fischer, S. C. (2018). Intelligence artificielle: les ambitions de la Chine. *Politique de sécurité: analyses du CSS*, *220*.

Fouse, S., Cross, S., & Lapin, Z. (2020). DARPA's impact on artificial intelligence. *Ai Magazine*, *41*(2), 3-8.

Fuhrer 1, C. (2023). Intelligence Artificielle: que dit la recherche récente? Une approche combinée bibliométrique et textuelle. *Management & Avenir*, (5), 89-111.

Future of Life Institute. (2017).

Ganascia, J. G. (2017). *Le Mythe de la Singularité. Faut-il craindre l'intelligence artificielle?*. Média Diffusion.

Gary, A. (2021). Intelligence artificielle et armées françaises: une technologie du présent à mettre en œuvre immédiatement. *Revue Défense Nationale*, (HS4), 200-213.

Gauthy Choc, L. (2023). État des lieux et analyse de l'apport potentiel du Cloud Computing et des techniques de Re connaissance Optique de Caractères dans le métier d'auditeur.

Giugnatico, I. (2020). Construction d'une épistémologie critique pour la recherche interdisciplinaire à des fins d'action (intelligence artificielle et médecine personnalisée).

Godin, V. (2021). Singularité militaire et digitalisation: les défis de l'intelligence artificielle dans l'optimisation des ressources humaines. *Revue Defense Nationale*, (HS4), 227-240.

Goffi, E. (2022). L'institutionnalisation des normes éthiques dans le domaine de l'intelligence artificielle: une construction sociale à dépasser. *Revue Ethique et Numérique*, *1*(1), 18-35.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Gornet, M., & Maxwell, W. (2023, July). Normes techniques et éthique de l'IA. In *Conférence Nationale en Intelligence Artificielle (CNIA)*.

Goutefangea, P. (2017). Alan Turing et l'intelligence artificielle: le «jeu de l'imitation» et «l'IA forte».

Guérin, A. (2021). Relancer la géographie française avec l'intelligence artificielle. *Revue Défense Nationale*, (7), 78-82.

Hadjitchoneva, J. (2020). L'intelligence artificielle au service de la prise de décisions plus efficace. *Pour une recherche économique efficace*, *149*.

HAMDANI, F., MONTICOLO, D., & BOLY, V. Etude de l'apport de l'Intelligence Artificielle pour l'innovation de produit.

High-Level Expert Group on AI. (2019). Ethics Guidelines for Trustworthy AI.

Hoadley, D. S., & Lucas, N. J. (2020). Artificial Intelligence and National Security. Congressional Research Service.

Hodges, A. (2002). Alan turing.

Holzinger, A. (2021, September). The next frontier: AI we can really trust. In *Joint European conference on machine learning and knowledge discovery in databases* (pp. 427-440). Cham: Springer International Publishing.

Horowitz, M. C. (2018). The Promise and Peril of Military Applications of Artificial Intelligence. Bulletin of the Atomic Scientists

Horowitz, M. C. (2019). The Ethics & Morality of Robotic Warfare: Assessing the Debate over Autonomous Weapons. Daedalus, 148(2), 25-36.

Idoux, P. (2022). LE TEMPS DE LA REGULATION. *Le temps en droit administratif*.

IEEE (2020). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems.

IIHL (2020). Artificial Intelligence and International Humanitarian Law.

INRIA. (2021). Partenariat pour une IA éthique et fiable.

Institut Montaigne. (2021). Réussir l'Intelligence Artificielle: Bilan et propositions pour la France. Rapport de l'Institut Montaigne.

Institute for Ethics and Emerging Technologies (IEET) (2021). The Need for Transparency in Military AI.

Intissar, S., & Rabeb, C. (2015). Étapes à suivre dans une analyse qualitative de données selon trois méthodes d'analyse: la théorisation ancrée de Strauss et Corbin, la méthode d'analyse qualitative de Miles et Huberman et l'analyse thématique de Paillé et Mucchielli, une revue de la littérature. *Revue francophone internationale de recherche infirmière*, *1*(3), 161-168.

Jean-Christophe, N. O. Ë. L. (2018). Intelligence artificielle: vers une nouvelle révolution militaire?. *Focus stratégique*, (84).

Khalaf, B. A., Mostafa, S. A., Mustapha, A., Mohammed, M. A., & Abduallah, W. M. (2019). Comprehensive review of artificial intelligence and statistical approaches in distributed denial of service attack and defense methods. *IEEE Access*, *7*, 51691-51713.

Kingston, J. (2018). Artificial intelligence and the future of warfare. Chatham House.

Krichen, M. (2023). Renforcer la sécurité des contrats intelligents grâce à la puissance de l'intelligence artificielle.

Langevin, B., & Begeot, F. (1995). Comparabilité et synthèse des données européennes: l'expérience d'Eurostat. *Josiane DUCHÊNE et Guillaume WUNSCH (éd.), Collecte et comparabilité des données démo-sociales en Europe*, 43-64.

Layton, P. (2021). Fighting Artificial Intelligence Battles: Operational Concepts for Future AI-Enabled Wars. *Network*, *4*(20), 1-100.

Leroy, C. (2023). Éthique de l'intelligence artificielle et transhumanisme: l'immortalité fait-elle la vie? 1. *Le Philosophoire*, *60*(2), 113-132.

Ma, A. (2018). L'intelligence artificielle en Chine: un état des lieux.
Mangelsdorf, A., Wittenbrink, N., & Gabriel, P. (2022). Regulierung und Zertifizierung von KI in der Industrie: Ziele, Kriterien und Herausforderungen. In *Digitalisierung souverän gestalten II: Handlungsspielräume in digitalen Wertschöpfungsnetzwerken* (pp. 110-119). Springer Berlin Heidelberg.

Martin, A. (2022, June). L'intelligence artificielle peut-elle être saisie par le droit de l'Union européenne?. In *Conférence Nationale en Intelligence Artificielle 2022 (CNIA 2022)*.
Marty, F. (2017). Algorithmes de prix, intelligence artificielle et équilibres collusifs 1. *Revue internationale de droit économique*, *31*(2), 83-116.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1956). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.

Meneceur, Y. (2020). La critique de la technique: clé du développement de l'intelligence artificielle?. *Les Temps Électriques en septembre*, *2020*, 42.

Mercier, C. (2023, October). Optimisation pédagogique grâce à l'Intelligence Artificielle: Un avenir éducatif innovant. In *Journée Pédagogique Régionale*.

Meunier, F., & Bombal, S. (2018). Prospective sur l'intelligence artificielle au profit de la cyberdéfense. *Revue Défense Nationale*, (4), 72-77.

Miailhe, N. (2018). Géopolitique de l'Intelligence artificielle: le retour des empires?. *Politique étrangère*, (3), 105-117.

Michaud, T. Vers un filtre éthique pour les intelligences artificielles chargées de la vente?.

Ministère de l'Économie, des Finances et de la Relance. (2021). Stratégie nationale pour l'intelligence artificielle.

Ministère des Armées. (2021). Confiance.AI: Le programme d'intelligence artificielle du ministère des Armées.

Mitchell, D., Blanche, J., Harper, S., Lim, T., Gupta, R., Zaki, O., ... & Flynn, D. (2022). A review: Challenges and opportunities for artificial intelligence and robotics in the offshore wind sector. *Energy and AI*, *8*, 100146.

Mori, S. (2018). US defense innovation and artificial intelligence. *Asia-Pacific Review*, *25*(2), 16-44.

National Institute of Standards and Technology. (2020). AI and Machine Learning.

NATO. (2020). Science & Technology Trends 2020-2040: Exploring the S&T Edge.

Nicolas, N. (2023). Kant et la morale des machines: programmer la «bonté» dans l'ère des modèles de langage. *Management & Datascience*, *7*(4).

OECD. (2021). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments.
OTAN (2019). Implications of Artificial Intelligence for Cybersecurity.

Parly, F. (2019). Avant-propos–Intelligence artificielle et défense. *Revue défense nationale*, (5), 9-17.

Perez, L. (2022). Intelligence artificielle de défense et approche non cinétique des conflits. In *Annuaire français de relations internationales* (pp. 33-49). Éditions Panthéon-Assas.

Petel, A. (2022). Quelle réglementation européenne sur l'intelligence artificielle?. *I2D–Information, données & documents*, (1), 22-28.
Pew Research Center (2020). Public Attitudes Toward Computer Algorithms.

Pires, A. (1997). Échantillonnage et recherche qualitative: essai théorique et méthodologique. *La recherche qualitative. Enjeux épistémologiques et méthodologiques*, *169*, 113.

Putzer, H. J., Rueß, H., & Koch, J. (2021). Trustworthy AI-based Systems with VDE-AR-E 2842-61. *Embedded World*.

Qiu, S., Liu, Q., Zhou, S., & Wu, C. (2019). Review of artificial intelligence adversarial attack and defense technologies. *Applied Sciences*, *9*(5), 909.

RAND Corporation (2020). Trust in Autonomous Systems
RARHOUI, K. (2023). Droit de l'Intelligence Artificielle et Administration publique. *Revue Internationale du Chercheur*, *4*(4).

Roland, A., & Shiman, P. (2002). Strategic Computing: DARPA and the Quest for Machine Intelligence, 1983-1993. MIT Press.

Romero, M., Heiser, L., & Lepage, A. (2023). Enseigner et apprendre à l'ère de l'intelligence artificielle: acculturation, intégration et usages créatifs de l'IA en éducation: livre blanc.

Sadek, D. (2019). De l'intelligence artificielle appliquée à la défense. *Revue Défense Nationale*, (5), 103-106.

Salomé, S., & Monfort, E. (2023). Révolution numérique et âgisme: les enjeux éthiques de l'intelligence artificielle pour les personnes âgées. *NPG Neurologie-Psychiatrie-Gériatrie*, *23*(138), 383-387.

Sarker, I. H., Furhad, M. H., & Nowrozy, R. (2021). Ai-driven cybersecurity: an overview, security intelligence modeling and research directions. *SN Computer Science*, *2*(3), 173.

Scharre, P. (2018). Army of None: Autonomous Weapons and the Future of War. W. W. Norton & Company.

Scharre, P. (2018). Army of None: Autonomous Weapons and the Future of War. W. W. Norton & Company.

Schmitt, M. N., & Thurnher, J. S. (2013). "Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict. Harvard National Security Journal, 4, 231-281.

Secrétariat général pour l'investissement (SGPI). (2021). Bilan et perspectives du programme Confiance.AI. Rapport officiel.
SEDJARI, A. IMPACT DU NUMÉRIQUE ET DE L'INTELLIGENCE ARTIFICIELLE SUR LES TRANSFORMATIONS DES GOUVERNANCES PUBLIQUES.

Senneville-Robert, M. (2021). *Reconfiguration des Liens de Collaboration Entre Acteurs Industriels et Universitaires de la Recherche en Intelligence Artificielle à Montréal et à Toronto* (Doctoral dissertation, Institut National de la Recherche Scientifique (Canada)).

Shiohira, K. (2021). Comprendre l'impact de l'intelligence artificielle sur le développement des compétences.

Soriano, J. (2018). *Les enjeux de l'intégration de solutions d'intelligence artificielle au sein d'OBNL* (Doctoral dissertation, PhD thesis).

Soulez, M., à la Cour d'appel, A., & de Paris, L. A. B. A. (2018). Questions juridiques au sujet de l'intelligence artificielle. *Enjeux numériques*, (1), 81-85.

Srivastava, A., Rehm, G., & Schneider, J. M. (2017, August). DFKI-DKT at SemEval-2017 Task 8: Rumour detection and classification using cascading heuristics. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)* (pp. 486-490).

Taddeo, M., & Floridi, L. (2018). Regulate artificial intelligence to avert cyber arms race. Nature.

Tadjdeh, Y. (2019). DARPA's 'AI next'program bearing fruit. *National Defense*, *104*(788), 8-8.

The State Council of the People's Republic of China. (2017). Next Generation Artificial Intelligence Development Plan.

Thibout, C. (2019). La compétition mondiale de l'intelligence artificielle. *Pouvoirs*, (3), 131-142.

Thiry, O., & Strowel, A. Le profilage à l'aide de techniques algorithmiques par les autorités publiques: la transparence et l'explicabilité, des garanties impossibles à assurer?.

Truong, T. C., Diep, Q. B., & Zelinka, I. (2020). Artificial intelligence in the cyber domain: Offense and defense. *Symmetry*, *12*(3), 410.

Turing, A. M. (1950). Computing Machinery and Intelligence. Mind, LIX(236), 433-460.

Turing, A. M., Girard, J. Y., Basch, J., & Blanchard, P. (1995). *La machine de Turing* (pp. 47-102). Editions du seuil.

Tyugu, E. (2011, June). Artificial intelligence in cyber defense. In *2011 3rd International conference on cyber conflict* (pp. 1-11). IEEE.

UAZRAOUI, N. (2014). *Application des Techniques de l'Intelligence Artificielle aux Problèmes de Gestion des Risques Industriels* (Doctoral dissertation, Université de Batna 2).

UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence.

UNIDIR. (2017). The Weaponization of Increasingly Autonomous Technologies: Considering Ethics and Social Values. United Nations Institute for Disarmament Research.

UNIDIR. (2017). The Weaponization of Increasingly Autonomous Technologies: Considering Ethics and Social Values. United Nations Institute for Disarmament Research.

United Nations Office for Disarmament Affairs (UNODA). (2019). Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapon Systems.

UNODA (United Nations Office for Disarmament Affairs). (2018). Lethal Autonomous Weapons Systems. United Nations.

Ventre, D. (2020). *Intelligence artificielle, cybersécurité et cyberdéfense* (Vol. 2). ISTE Group.

Villani, C. (2018). Donner un sens à l'intelligence artificielle: Pour une stratégie nationale et européenne. Rapport au Premier ministre.

Vitanage, D., Doolan, C., Maunsell, L., Cameron, B., Wang, Y., & Li, Z. (2018). SUCCESS IN DATA ANALYTICS-SYDNEY WATER AND DATA61 COLLABORATION. *Water e-Journal*.

Voarino, N. (2020). Systèmes d'intelligence artificielle et santé: les enjeux d'une innovation responsable.

Wiener, N. (1971). Cybernétique et société (1950). *Paris: Union Générale d'Editions*.
Wiener, N. (2014). *La Cybernétique. Information et régulation dans le vivant et la machine: Information et régulation dans le vivant et la machine*. Média Diffusion.

Wright, R. (2021). Sommaire Réglementer l'intelligence artificielle-Enjeux et choix essentiels.

Xinhua News Agency. (2018). China Focus: China's AI Development Plan Takes Shape.

Youssef, E. R., LEMQEDDEM, H. A., & EZZAHIRI, M. (2022). La posture épistémologique en science de gestion: quelle revue de littérature?. *Revue Internationale des Sciences de Gestion*, *5*(1).

Zeng, Y., Lu, E., & Huangfu, C. (2018). Linking Artificial Intelligence Principles. AI & Society.

Zhang, Y., Dai, Z., Zhang, L., Wang, Z., Chen, L., & Zhou, Y. (2020, December). Application of artificial intelligence in military: From projects view. In *2020 6th International Conference on Big Data and Information Analytics (BigDIA)* (pp. 113-116). IEEE.
Zouinar, M. (2020). Évolutions de l'Intelligence Artificielle: quels enjeux pour l'activité humaine et la relation Humain-Machine au travail?. *Activités*, (17-1).

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs.