Nations Unies A/80/78

Assemblée générale

Distr. générale
5 juin 2025
Français
Original : anglais/chinois/espagnol/français/russe

Quatre-vingtième session
Point 101 de la liste préliminaire*
Désarmement général et complet

L'intelligence artificielle dans le domaine militaire et ses conséquences pour la paix et la sécurité internationales

Rapport du Secrétaire général

Résumé

Synthèse des communications reçues des États Membres et des États observateurs comme suite à la résolution 79/239 de l'Assemblée générale, sans préjudice de la position de chacun d'eux sur la question, le présent rapport fait le point sur les possibilités et difficultés liées à l'utilisation de l'intelligence artificielle dans le domaine militaire, les propositions normatives existantes et émergentes, les initiatives relatives à l'intelligence artificielle dans le domaine militaire, les facteurs à prendre en considération pour les prochaines étapes, ainsi que les observations et conclusions du Secrétaire général.

* A/80/50.

25-06526 (F) 190625 050825

Table des matières

I. Introduction

1. Au paragraphe 7 de la résolution 79/239 sur l'intelligence artificielle dans le domaine militaire et ses conséquences pour la paix et la sécurité internationales, l'Assemblée générale a prié le Secrétaire général de solliciter les vues des États Membres et des États observateurs sur les possibilités et les difficultés que l'application de l'intelligence artificielle dans le domaine militaire présente pour la paix et la sécurité internationales, en particulier dans des domaines autres que les systèmes d'armes létaux autonomes, et de lui présenter, à sa quatre-vingtième session, un rapport de fond résumant ces vues et répertoriant les propositions normatives existantes et nouvelles, assorti d'une annexe contenant ces vues, dans la perspective de futurs débats entre les États. Au paragraphe 8 de la même résolution, l'Assemblée a également prié le Secrétaire général d'inviter les organisations internationales et régionales, le Comité international de la Croix-Rouge, la société civile, la communauté scientifique et les professionnels du secteur à faire part de leurs points de vue, lesquels seront inclus dans l'annexe du rapport susmentionné dans la langue de l'original. Le présent rapport fait suite à ces demandes.

2. Le 12 février 2025, le Bureau des affaires de désarmement a adressé à tous les États Membres et États observateurs une note verbale pour appeler leur attention sur le paragraphe 7 de la résolution 79/239 et a cherché à obtenir leurs vues à ce propos. Des notes verbales et des lettres ont aussi été envoyées aux entités visées au paragraphe 8 de ladite résolution, pour attirer leur attention sur ledit paragraphe et solliciter leur avis sur la question. On trouvera dans les annexes au présent rapport le texte des communications reçues au 11 avril 2025. Toute contribution reçue après cette date sera publiée sur le site Web du Bureau des affaires de désarmement dans la langue de l'original.

3. Les sections II à VI du présent rapport font la synthèse des communications reçues des États Membres et des États observateurs, sans préjudice de la position de chacun d'eux sur la question. On trouvera à la section VII les observations et conclusions du Secrétaire général.

II. Contexte

4. Les États ont évoqué les progrès rapides de la science et de la technologie en général, et de l'intelligence artificielle (IA) en particulier, en soulignant leurs répercussions considérables sur la société. Plus précisément, les États ont constaté que l'IA était susceptible de transformer tous les aspects des affaires militaires et d'avoir des retombées majeures sur la paix et la sécurité internationales.

5. Plusieurs États ont évoqué les applications actuelles de l'IA dans le domaine militaire, ainsi que les efforts qu'ils avaient déployés pour utiliser l'IA dans les opérations de défense. Conscients de l'importance des débats sur les systèmes d'armes létaux autonomes, les États ont noté que la question de l'IA dans le domaine militaire dépassait ce cadre et englobait un plus grand nombre de capacités.

III. Possibilités et difficultés

6. Il a été noté que l'IA créait à la fois des possibilités et des difficultés, qui devaient être abordées de manière réaliste. Les États ont constaté qu'à l'heure actuelle, le rythme de développement de l'IA ne permettait pas de prévoir l'ensemble de ces possibilités et difficultés, et suggéré de ne pas stigmatiser la technologie elle-même.

A. Possibilités

7. Parmi les avantages de l'IA, sa rapidité a été citée comme un avantage majeur, notamment dans l'analyse des informations et la prise de décision. Son envergure est un autre des avantages cités, l'IA pouvant jouer le rôle de « multiplicateur de force ». Plusieurs États ont évoqué le potentiel de l'IA pour ce qui est de l'amélioration de l'efficacité, de l'exactitude et de la précision, et de la réduction, par conséquent, du risque d'erreur (par comparaison avec les humains). Les autres caractéristiques relevées sont la fiabilité, la sécurité et la robustesse.

Applications

8. Plusieurs États ont évoqué les applications de l'IA dans les domaines du renseignement, de la surveillance et de la reconnaissance : elle pourrait être utilisée pour analyser efficacement de vastes jeux de données, faciliter la détection des menaces, améliorer la perception de la situation et permettre des opérations plus précises. Il a été noté que ces mêmes caractéristiques permettaient d'utiliser l'IA pour contribuer à la prise de décision ainsi qu'au commandement et au contrôle, ce qui pourrait conduire à des opérations plus précises, à la réduction des risques pour les civils et à une meilleure protection des biens de caractère civil. Toutefois, il a aussi été souligné que les outils utilisant l'IA ne sauraient remplacer la prise de décision par des humains.

9. Plusieurs États ont indiqué que l'IA pourrait être intégrée à des systèmes sans pilote. Il a été noté que l'IA pouvait améliorer la coordination et la communication entre les acteurs militaires d'une part et entre les acteurs militaires et d'autres acteurs d'autre part (ceux qui apportent une aide humanitaire, par exemple). D'une manière générale, on a fait remarquer que l'IA pouvait réduire la charge créée par les tâches routinières ou répétitives et augmenter les capacités humaines dans les tâches complexes.

10. Selon certains États, l'IA pourrait être utilisée pour renforcer la sécurité des technologies de l'information et des communications en détectant les intrusions ou d'autres activités malveillantes, ainsi que pour protéger les infrastructures critiques. Il a été noté que le recours à l'IA pourrait favoriser la détection de contenus générés par d'autres IA et utilisés à des fins de mésinformation et de désinformation, et la mise au jour des discours de haine, de la propagande ou de l'évolution de l'opinion publique.

11. D'autres applications de l'IA qui ne sont pas directement liées au combat ont été citées, notamment l'optimisation de la logistique, la maintenance prédictive, les achats, l'allocation des ressources, l'administration, la simulation et la formation.

Paix et sécurité internationales

12. Plusieurs États ont estimé que l'IA pouvait contribuer au maintien de la paix et de la sécurité internationales : la perception de la situation grâce à l'IA étant susceptible, par exemple, d'aider à atténuer les risques et de favoriser la désescalade des conflits. Il a été noté que l'utilisation de l'IA pourrait réduire les risques pour le personnel militaire, par exemple en remplaçant les humains dans certaines tâches dangereuses telles que l'élimination d'engins non explosés, ou en soutenant les opérations de recherche et de sauvetage menées dans des endroits éloignés.

13. Il a été suggéré que l'IA pourrait améliorer l'application du droit international humanitaire, en particulier ses principes fondamentaux que sont le principe de distinction, le principe de proportionnalité et l'obligation de prendre des précautions, ainsi que la protection des civils et des biens de caractère civil. À cet égard, plusieurs États ont noté que l'IA pouvait améliorer la perception de la situation en général, et

la compréhension de l'environnement civil en particulier, et accroître la précision tout en réduisant le risque d'erreur humaine. Il a également été constaté que l'IA pouvait faciliter les enquêtes sur les victimes civiles et garantir ainsi que les responsables rendent des comptes.

14. Plusieurs États ont suggéré que l'IA pourrait contribuer au suivi et à la vérification de l'application des accords de désarmement, de non-prolifération et de maîtrise des armements. Il a été fait référence au potentiel de l'IA aux fins de l'appui aux missions de maintien de la paix, notamment par la facilitation de la planification, de la logistique et de la surveillance du cessez-le-feu. D'autres applications connexes de l'IA ont été citées, notamment la sécurité des frontières, la lutte contre le terrorisme, la détection des programmes d'armement illégaux et l'optimisation de l'aide humanitaire et des secours en cas de catastrophe.

B. Difficultés

15. Plusieurs États ont noté que l'évolution rapide des technologies émergentes en général et de l'intelligence artificielle en particulier posait des problèmes pour la paix et la sécurité internationales. Il est important de comprendre ces problèmes mais, à l'heure actuelle, il n'est pas possible de tous les prévoir.

16. En ce qui concerne l'intelligence artificielle, les préoccupations suivantes ont été mises en avant :

- accélération de la boucle Observation, Orientation, Décision, Action, qui comprime le temps disponible pour la prise de décision ;
- autonomie croissante et perte de contrôle humain, en particulier dans le contexte de l'emploi de la force ;
- risque d'utilisation abusive ou malveillante ;
- confiance excessive des humains dans les applications utilisant l'IA ;
- renforcement des asymétries technologiques entre les États.

Paix et sécurité internationales

17. Plusieurs États ont noté que l'intégration de l'IA dans le domaine militaire pourrait poser des problèmes pour la paix et la sécurité internationales. L'utilisation de l'IA pourrait accroître le risque de malentendu, d'erreur d'appréciation et d'escalade involontaire, notamment en raison de la vitesse et de l'échelle accrues des opérations soutenues par l'IA ou en raison de défaillances techniques. Ces facteurs pourraient aussi abaisser le seuil d'emploi de la force. Plusieurs États ont fait part de leur préoccupation quant à l'émergence d'une course aux armements dans ce domaine. Il a été suggéré que l'utilisation de l'IA pourrait déplacer l'équilibre de la défensive vers l'offensive, et que des disparités croissantes entre les États pourraient conduire à une instabilité accrue, ce qui mettrait en péril la paix et la sécurité internationales.

18. Plusieurs États se sont inquiétés de l'effet potentiellement déstabilisateur de la prolifération des capacités d'IA, y compris pour les acteurs non étatiques. Il a été noté qu'il n'existe actuellement aucun cadre multilatéral visant à contrôler la prolifération des armes qui intègrent des capacités d'IA.

Considérations technologiques

19. Les États ont pris en compte les risques découlant de considérations technologiques, notamment :

- les défaillances techniques et les dysfonctionnements ;
- les défauts de conception ;
- les comportements inattendus, qui s'écartent des paramètres de conception ;
- la vulnérabilité aux cyberattaques et à l'empoisonnement des données ;
- les biais des algorithmes et des données, y compris les préjugés liés au genre ;
- les biais d'automatisation, résultant d'opérateurs humains ne disposant pas d'une formation suffisante ;
- les inquiétudes en matière de protection de la vie privée découlant de la collecte et du traitement de grandes quantités de données à caractère personnel utilisées pour former des modèles d'IA ;
- les problèmes causés par des modèles d'IA mal entraînés ;
- les problèmes découlant de procédures d'essai, d'évaluation, de validation et de vérification insuffisantes ;
- les erreurs de sélection des cibles ;
- la consommation d'énergie excessive ;
- la dépendance excessive envers des fournisseurs externes.

20. Plusieurs États se sont alarmés de la transparence et de l'explicabilité pour ce qui concerne les capacités complexes de l'IA, souvent qualifiées de « boîtes noires ». Des préoccupations ont aussi été exprimées quant à l'utilisation d'applications civiles de l'IA, telles que l'IA générative, qui pourraient ajouter de la complexité et de l'incertitude à une situation de conflit. Par ailleurs, plusieurs États ont fait part de leurs inquiétudes quant à la convergence de l'IA et d'autres technologies.

Considérations juridiques et humanitaires

21. Plusieurs États ont fait remarquer que l'IA créait des enjeux en matière de respect du droit international, en particulier du droit international humanitaire et du droit international des droits humains. L'utilisation de l'IA peut conduire à un emploi de la force indiscriminé et soulève des questions de responsabilité en cas d'actes illégaux ou répréhensibles. Parmi les questions connexes abordées figurent la protection des civils et des infrastructures civiles, ainsi que la possibilité d'accroître l'intensité et la létalité des conflits pour les combattants.

22. Plusieurs États ont fait part de leurs préoccupations éthiques, et observé que l'utilisation de l'IA pourrait réduire le champ d'application de la compassion, du raisonnement moral et du jugement humain.

Domaines potentiels de mauvaise utilisation

23. Plusieurs États ont noté que des acteurs étatiques et non étatiques pouvaient avoir recours à l'IA pour conduire des cyberattaques, y compris des attaques ciblant des infrastructures critiques. L'IA pourrait aussi être utilisée pour des campagnes de mésinformation et de désinformation, et notamment pour la production de fausses informations et d'hypertrucages (deepfakes) et pour la diffusion de telles informations par des robots pilotés par l'IA. Le recours à la mésinformation et à la désinformation, pour influencer des élections par exemple, pourrait avoir un effet déstabilisateur.

Armes de destruction massive

24. Plusieurs États ont souligné qu'il importait de maintenir un contrôle humain sur les armes nucléaires et leurs vecteurs et se sont inquiétés de la possibilité d'intégrer l'IA dans les systèmes de commandement, de contrôle et de communication nucléaires. L'engagement pris par certains États dotés d'armes nucléaires de maintenir le contrôle humain et l'intervention humaine à toutes les étapes essentielles de la formation de décisions souveraines concernant l'emploi des armes nucléaires et de leur exécution a notamment été cité, et on a attiré l'attention sur les répercussions potentielles pour la stabilité stratégique et l'escalade.

25. Plusieurs États se sont inquiétés du fait que l'IA pourrait faciliter la prolifération des armes de destruction massive, y compris au profit d'acteurs non étatiques. Dans ce contexte, le fait que l'IA puisse être utilisée pour développer et produire des armes biologiques s'avère particulièrement préoccupant. Il a été souligné que, comme suite aux dispositions des traités existants, l'IA ne doit pas être utilisée à cette fin. Il a également été dit que l'IA pourrait servir à freiner la prolifération des armes de destruction massive.

IV. Propositions normatives existantes et nouvelles

26. Plusieurs États ont indiqué que l'IA devait être employée à des fins pacifiques, notamment pour le règlement pacifique des différends. Les États ont aussi mis en avant l'importance de traiter et d'atténuer les risques créés par l'IA dans le domaine militaire, certains notant que les problèmes créés par l'IA militaire devraient être abordés collectivement.

27. En ce qui concerne l'IA dans le domaine militaire, les États ont demandé que l'approche adoptée soit :

- souple, équilibrée, réaliste et progressive, et donc capable de s'adapter aux progrès technologiques ;
- prudente ;
- axée sur l'ensemble du cycle de vie de l'IA, notamment la préconception, la conception, le développement, l'évaluation, la mise à l'essai, le déploiement, l'utilisation, la vente, l'achat, l'exploitation et la mise au rebut ;
- fondée sur les applications et l'utilisation de l'IA, plutôt que sur la technologie elle-même ;
- et qu'elle tienne compte des obligations existantes.

Il a été suggéré que les efforts déployés dans ce domaine devraient clairement faire la distinction entre les utilisations létales et non létales.

Considérations juridiques

28. Les États ont rappelé la résolution 79/239, dans laquelle l'Assemblée générale a affirmé que le droit international, notamment la Charte des Nations Unies, le droit international humanitaire et le droit international des droits humains, s'appliquait aux questions qu'il régit et qui se posent à tous les stades du cycle de vie de l'intelligence artificielle, y compris des systèmes basés sur l'intelligence artificielle, dans le domaine militaire. Il a été noté que le droit international en général, et le droit international humanitaire en particulier, n'interdisait pas catégoriquement l'utilisation de capacités d'IA.

29. Les États ont affirmé qu'ils respectaient le droit international dans leur utilisation de l'IA dans le domaine militaire. Il a été suggéré que le respect des obligations juridiques, en particulier celles qui découlent du droit international, doit être une considération essentielle dans la gouvernance, la conception et le déploiement de l'IA dans le domaine militaire. Il a aussi été dit que l'IA devrait être conçue pour renforcer le respect du droit international humanitaire. Plusieurs États ont souligné l'importance de procéder à des examens juridiques des nouvelles armes, moyens ou méthodes de guerre à cet égard.

30. Plusieurs États ont mis en exergue l'importance de tenir compte des considérations éthiques, en plus des cadres juridiques.

Considérations relatives à la paix et à la sécurité internationales

31. Plusieurs États ont noté que l'IA utilisée dans le domaine militaire devrait renforcer la paix et la sécurité internationales et être employée de manière à ne pas conduire à l'instabilité ou à l'escalade. On a fait valoir que les États devraient s'abstenir de rechercher un avantage militaire absolu grâce à l'IA et devraient veiller à ce que cette technologie ne devienne pas un outil servant à lancer une invasion et à poursuivre des visées hégémoniques.

32. Plusieurs États ont rappelé que l'IA ne devait pas porter atteinte aux accords de désarmement, de non-prolifération et de maîtrise des armements existants. Certains ont demandé que des efforts soient déployés afin d'empêcher la prolifération de la technologie de l'IA au profit d'acteurs non étatiques. Il a été souligné qu'il importait d'éviter les mécanismes de surveillance internationale arbitraires ou les contrôles à l'exportation discriminatoires.

Utilisation responsable de l'intelligence artificielle dans le domaine militaire

33. Plusieurs États ont estimé que l'IA devait être utilisée de manière responsable tout au long de son cycle de vie. Il a aussi été dit que le concept de responsabilité devrait être lié à la légalité et à l'obligation de rendre des comptes.

34. Plusieurs États ont souligné l'importance d'une approche centrée sur l'être humain pour ce qui concerne l'IA, et du contrôle et de la responsabilité exercés par les humains à tout moment. Il a été fait référence à l'importance de concepts tels que le « contrôle humain et le jugement adaptés au contexte » et le « contrôle humain significatif ». En revanche, selon d'autres États, ces concepts ne sont pas suffisamment définis. Il a été signalé que l'utilisation du concept de « contrôle humain significatif » pourrait entraver la recherche légitime ou restreindre indûment l'utilisation de l'IA dans le domaine militaire.

35. Les États ont fait ressortir qu'il importe de garantir la responsabilité humaine et l'obligation de rendre des comptes, y compris en présence d'un contrôle humain et d'une chaîne de commandement responsable, comme le prévoit le droit international.

Considérations technologiques

36. Les États ont examiné les principes de gouvernance d'un point de vue technologique, notamment les suivants :

- la sûreté (aux fins de la robustesse des systèmes d'IA face aux menaces extérieures) ;
- la sécurité (notamment par l'incorporation de garde-fous visant à minimiser les dommages) ;
- la fiabilité (pour éviter les conséquences involontaires et les dysfonctionnements) ;

- des limites et des contraintes opérationnelles claires (pour éviter les comportements non intentionnels) ;
- des cas d'utilisation bien définis ;
- la gouvernabilité (pour assurer des interactions adéquates entre l'humain et la machine et l'atténuation des biais) ;
- l'équité et la justice ;
- la protection de la vie privée ;
- l'explicabilité, l'intelligibilité et la traçabilité ;
- la transparence.

37. L'utilisation de données de formation permettant de respecter pleinement le droit international a été soulignée. Plusieurs États ont affirmé avec insistance l'importance des essais tout au long du cycle de vie pour détecter les erreurs et garantir la fiabilité. Les États ont aussi mis l'accent sur le fait qu'une formation adéquate du personnel travaillant avec l'IA était nécessaire pour atténuer les risques et garantir le respect du droit humanitaire international. L'importance du contrôle des performances des systèmes tout au long de leur cycle de vie et des mesures visant à désactiver les systèmes en toute sécurité au moment de leur mise hors service a été signalée.

V. Initiatives relatives à l'intelligence artificielle dans le domaine militaire

Instances internationales

38. Plusieurs États ont fait état des débats en cours à l'Organisation des Nations Unies, du Pacte pour l'avenir (résolution 79/1 de l'Assemblée générale) et de son annexe, le Pacte numérique mondial, et de la résolution de l'Assemblée générale sur l'intelligence artificielle dans le domaine militaire et ses conséquences pour la paix et la sécurité internationales (résolution 79/239). La réunion du Conseil de sécurité organisée selon la formule Arria sur l'utilisation d'une intelligence artificielle sûre, inclusive et digne de confiance pour le maintien de la paix et de la sécurité internationales, qui s'est tenue le 4 avril 2025, a également été citée.

39. Les États ont fait référence aux discussions multilatérales menées sur des sujets liés à l'IA dans le domaine militaire, telles que celles qui ont eu lieu à la Commission du désarmement au titre du point de l'ordre du jour intitulé « Recommandations visant à promouvoir une communauté de vues sur les questions relatives aux technologies émergentes dans le contexte de la sécurité internationale », les travaux du Groupe d'experts gouvernementaux sur les technologies émergentes dans le domaine des systèmes d'armes létaux autonomes établi dans le cadre de la Convention sur l'interdiction ou la limitation de l'emploi de certaines armes classiques qui peuvent être considérées comme produisant des effets traumatiques excessifs ou comme frappant sans discrimination, et les résolutions de l'Assemblée générale sur les systèmes d'armes létaux autonomes (résolutions 78/241 et 79/62).

40. Les États ont aussi mentionné leur participation aux activités liées à l'IA dans le domaine militaire organisées par le Bureau des affaires de désarmement et l'Institut des Nations Unies pour la recherche sur le désarmement.

Initiatives menées par l'État

41. Plusieurs États ont fait savoir qu'ils avaient pris des initiatives liées à l'IA dans le domaine militaire ou qu'ils avaient participé à de telles initiatives, par exemple :

- la Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles et le processus d'application qui a suivi ;
- l'initiative « Responsible Artificial Intelligence in the Military Domain », qui a vu la tenue de conférences au Royaume des Pays-Bas en 2023, au cours de laquelle un appel à l'action a été approuvé, et en République de Corée en 2024, au cours de laquelle un plan d'action a été approuvé. La Commission mondiale sur l'intelligence artificielle responsable dans le domaine militaire devrait publier un rapport avant la prochaine conférence, qui se tiendra en Espagne en 2025 ;
- le Sommet pour l'action sur l'intelligence artificielle, qui s'est tenu en France en 2025, au cours duquel a été adoptée la Déclaration de Paris sur le maintien du contrôle humain dans les systèmes d'armes dotés d'IA ;
- le Sommet sur la sécurité de l'IA qui s'est tenu au Royaume-Uni de Grande-Bretagne et d'Irlande du Nord en 2023, au cours duquel la Déclaration de Bletchley a été adoptée ;
- les travaux sur l'IA menés dans le cadre du Groupe des Sept ;
- le partenariat « AI Partnership for Defence » ;
- l'Initiative mondiale pour la gouvernance de l'intelligence artificielle proposée en 2023.

42. Il a été estimé que ces initiatives, bien qu'utiles, pourraient entraîner une fragmentation. Des inquiétudes ont aussi été exprimées quant au fait que les résultats de ces initiatives ne tenaient pas compte des points de vue de tous les États concernés et risquaient de compromettre l'inclusivité dans ce domaine.

Initiatives régionales

43. Les États ont noté l'importance des initiatives régionales pour ce qui est de favoriser des discussions inclusives et propres au contexte sur l'IA dans le domaine militaire. Les exemples à cet égard sont notamment les suivants :

- la déclaration conjointe sur la coopération en matière d'IA dans le secteur de la défense, adoptée lors de la retraite de la réunion des ministres de la défense de l'ASEAN en 2025 ;
- la seizième Conférence des Ministres de la défense des Amériques, tenue en 2024 et au cours de laquelle la Déclaration de Mendoza a été adoptée ;
- les activités menées dans le cadre de l'Organisation du Traité de l'Atlantique Nord, y compris sa Stratégie pour l'intelligence artificielle, révisée pour la dernière fois en 2024, et ses principes pour une utilisation responsable, élaborés en 2021 ;
- les consultations régionales menées en 2024 au Chili, au Kenya, au Royaume des Pays-Bas, à Singapour et en Türkiye dans le cadre de l'initiative sur l'intelligence artificielle responsable dans le domaine militaire (« Responsible Artificial Intelligence in the Military Domain »).

Initiatives nationales

44. Les États ont évoqué les activités menées à l'échelle nationale, qui portent entre autres sur la législation, les réglementations, les stratégies et les organes existants en matière d'IA, ainsi que les efforts déployés pour les développer.

VI. Prochaines étapes

45. Les États ont appelé à un dialogue sur l'IA dans le domaine militaire. Plusieurs États ont demandé que les répercussions sur la paix et la sécurité internationales de l'utilisation de l'IA dans le domaine militaire fassent l'objet d'une étude plus approfondie.

46. De nombreux États ont indiqué que la poursuite du dialogue devrait avoir pour objectif d'atténuer les risques posés par l'IA dans le domaine militaire. Il a été suggéré que le dialogue devrait viser l'instauration de cadres réglementaires ou de dispositifs de gouvernance. Plusieurs États ont demandé que des normes, des règles et des principes soient élaborés pour régir le cycle de vie de l'IA dans le domaine militaire. Si certains États ont exprimé leur préférence pour l'élaboration d'un cadre juridiquement contraignant, d'autres ont estimé que l'adoption de nouvelles mesures juridiques n'était pas nécessaire à l'heure actuelle. Il a également été dit que les normes, les règles et les principes pourraient constituer la base d'engagements juridiques à un stade ultérieur. Plusieurs États ont fait part de leur opposition au concept de normes, règles et principes de développement, de déploiement ou d'utilisation responsable, en soulignant que ce concept ne faisait pas l'objet d'un consensus. On a estimé que l'introduction prématurée de règlements n'était pas souhaitable.

47. L'importance d'éviter la duplication et la fragmentation de la gouvernance a été soulignée. Des États ont considéré qu'un débat sur la gouvernance devrait prendre en compte de manière équilibrée les considérations relevant du domaine humanitaire, de la sécurité et du développement. Les États ont indiqué qu'il importait d'éviter les restrictions qui entraveraient l'innovation légitime et le progrès technologique. Plusieurs États ont jugé que les utilisations pacifiques de l'IA, notamment par les pays en développement, ne devaient pas être entravées.

48. Il a été suggéré que toute approche relative à la gouvernance devrait tenir compte du fait que les États se trouvent à différents stades de l'intégration de l'IA dans les capacités militaires et que leurs environnements de sécurité sont différents. L'importance de la participation de tous les États aux discussions sur la gouvernance de l'IA dans le domaine militaire a été mise en exergue. De nombreux États ont estimé que les discussions futures devraient adopter une approche multipartite, incluant les organisations internationales et régionales, la société civile, la communauté scientifique et l'industrie. Il a toutefois été souligné que la prise de décision devait rester la prérogative exclusive des États.

49. Les États ont envisagé diverses priorités pour le dialogue futur sur l'IA dans le domaine militaire, notamment :

- veiller au respect du droit international, en particulier du droit international humanitaire ;
- protéger la dignité humaine et les droits humains ;
- rechercher une compréhension commune des définitions et de la terminologie ;
- envisager la transparence et les mesures de confiance ;
- examiner la question de l'autonomie dans l'emploi de la force ;

- étudier les systèmes d'intelligence artificielle qui soutiennent directement les opérations de combat ;
- veiller à ce que les mécanismes de gouvernance des données soient suffisants ;
- renforcer la coopération et l'assistance internationales ;
- soutenir le renforcement des capacités, notamment par le partage des connaissances, le transfert de technologie et l'échange de bonnes pratiques, afin de réduire la fracture numérique et le fossé en matière d'intelligence artificielle ;
- promouvoir un dialogue régional continu ;
- promouvoir une réglementation nationale, notamment pour que le secteur privé respecte le droit international.

50. Plusieurs États ont suggéré que l'examen des systèmes d'armes létaux autonomes devrait faire partie de tout débat portant sur l'IA dans le domaine militaire. Il a aussi été dit que les discussions actuellement menées sur ces systèmes étaient complémentaires des échanges relatifs à l'IA dans le domaine militaire. Plusieurs États ont rappelé leur position sur les systèmes d'armes létaux autonomes[1]. Si certains ont estimé que le Groupe d'experts gouvernementaux créé dans le cadre de la Convention sur certaines armes classiques était le meilleur espace pour discuter de l'IA dans le domaine militaire, d'autres ont déclaré que, compte tenu de son mandat et de sa composition non universelle, ce groupe n'était pas un espace approprié pour de telles discussions.

51. Plusieurs États ont demandé la tenue de discussions sur l'IA dans le domaine militaire dans le cadre des instances de l'Organisation des Nations Unies. Il a été suggéré que le présent rapport pourrait servir de base à ces échanges. Les États ont indiqué que les discussions futures devraient compléter les processus en cours, tels que le Groupe de travail à composition non limitée sur la sécurité du numérique et de son utilisation.

52. Plusieurs États ont estimé que les mécanismes de désarmement instaurés par l'Organisation des Nations Unies constituaient un espace efficace et inclusif et qu'ils devraient jouer un rôle central dans les discussions futures sur l'IA dans le domaine militaire. Il a été suggéré que la Conférence du désarmement pourrait examiner la question de l'IA, en particulier en ce qui concerne les armes nucléaires. Il a été estimé que la Première Commission de l'Assemblée générale pourrait aussi accueillir de tels échanges, et qu'elle pourrait demander au Secrétaire général de présenter des rapports réguliers faisant le point sur le développement technologique de l'IA dans le domaine militaire. Plusieurs États ont suggéré la tenue de débats dans le cadre de la Commission du désarmement.

53. Par ailleurs, il a été suggéré que des débats soient organisés dans le cadre du Conseil de sécurité.

54. Plusieurs États ont proposé la création d'un mécanisme propre, tel qu'un groupe de travail à composition non limitée, d'autres ont jugé qu'il n'était pas utile, pour l'instant, de créer un nouveau dispositif à l'ONU. Il a été estimé que tout dispositif de l'Organisation des Nations Unies sur cette question devrait être fondé sur le consensus.

[1] Pour plus de détails, voir A/79/88.

VII. Observations et conclusions du Secrétaire général

55. L'IA est susceptible d'influencer tous les aspects de notre vie. Lorsqu'elle est utilisée à des fins pacifiques, elle peut jouer un rôle important en facilitant la réalisation des engagements et des objectifs de développement, y compris les objectifs de développement durable.

56. Dans le domaine militaire, l'IA peut être utile aux armées qui l'emploient comme aux populations civiles, en augmentant la précision des opérations et en réduisant le risque d'erreur humaine. Dans le même temps, son utilisation dans le domaine militaire soulève des enjeux importants, dont le principal est le maintien de la responsabilité humaine et de l'obligation de rendre des comptes.

57. L'affirmation par l'Assemblée générale, dans sa résolution 79/239, que le droit international, notamment la Charte des Nations Unies, le droit international humanitaire et le droit international des droits humains, s'applique tout au long du cycle de vie des outils d'intelligence artificielle constitue un socle fondamental. Toutefois, d'importantes questions sur l'application de la loi doivent encore être résolues.

58. L'utilisation de l'IA militaire dans des situations impliquant l'emploi de la force nécessite une attention particulière. Malgré les avantages potentiels pour la protection des civils et des combattants, les utilisations de l'IA signalées dans les conflits actuels suscitent des inquiétudes quant au contrôle humain et au rôle joué par l'IA dans la facilitation des hostilités dans des zones densément peuplées. Les machines qui ont le pouvoir discrétionnaire de mettre fin à des vies humaines sont politiquement inacceptables et moralement révoltantes.

59. Les risques posés par les armes nucléaires ne disparaîtront pas tant que les armes elles-mêmes ne seront pas éliminées. En attendant la suppression totale des armes nucléaires, j'invite instamment tous les États qui possèdent ces armes à accepter que toute décision sur leur utilisation soit prise par des humains et non par des machines.

60. L'IA peut abaisser la barrière qui empêche les acteurs étatiques et non étatiques de mettre au point ou d'acquérir des armes chimiques et biologiques. Par conséquent, j'exhorte vivement les États à respecter pleinement les obligations qui leur incombent au titre des cadres de désarmement, de non-prolifération et de maîtrise des armements, à évaluer systématiquement les enjeux et les effets de l'IA pour ces cadres et à se préparer à y répondre.

61. L'intégration potentielle des applications civiles de l'IA dans le domaine militaire est une source d'inquiétude croissante. Le fait que les technologies utilisant l'IA sont, par nature, réutilisables soulève des enjeux en matière de surveillance, de transparence et de responsabilité. Je recommande fortement aux États d'examiner attentivement le flou qui entoure les évolutions des applications civiles de l'IA et leur utilisation potentielle dans le domaine militaire.

62. La mise en place de mécanismes de coopération supplémentaires en matière d'IA, en particulier aux niveaux régional et sous-régional, présente un grand intérêt. Les organisations régionales et sous-régionales sont particulièrement bien équipées pour élaborer et mettre en œuvre des mesures de transparence et de confiance afin d'atténuer les risques. J'encourage donc les États à envisager d'élaborer des mesures de transparence et de confiance aux niveaux régional et sous-régional, qui seraient adaptées aux caractéristiques et problèmes qui sont propres à l'IA.

63. Des discussions inclusives sur l'utilisation pacifique de l'IA et sa gouvernance au profit de l'humanité sont actuellement menées sous l'égide de l'Organisation des Nations Unies, en particulier dans le contexte de la mise en œuvre du Pacte numérique

mondial. Néanmoins, l'examen mené par les États Membres quant au recours à l'IA dans le domaine militaire s'est déroulé principalement en dehors de l'Organisation. La résolution 79/239 de l'Assemblée générale et le présent rapport constituent des premiers pas notables pour porter cette question importante devant l'Organisation. J'encourage les États à mener ces délibérations de façon inclusive et constructive, en vue de faire progresser la compréhension commune et de renforcer la coopération internationale pour atténuer les risques.

64. Les États sont encouragés à explorer les activités (y compris en matière de renforcement des capacités) qui permettraient d'assurer une véritable participation de tous les États aux travaux menés par l'Organisation sur ce sujet, ce qui est essentiel pour favoriser une compréhension partagée, développer des approches communes et atténuer les risques.

65. L'Assemblée générale sait parfaitement définir des mandats qui favorisent des débats ouverts sur les questions liées aux technologies émergentes et à la sécurité internationale, tout en encourageant la contribution des parties prenantes, notamment les organisations internationales et régionales, la société civile, la communauté scientifique et l'industrie. Cette approche multipartite est particulièrement importante dans le domaine de l'IA, où l'innovation est largement portée par le secteur privé et où une grande partie de l'expertise réside en dehors des gouvernements, dans les universités et la communauté scientifique.

66. **Je recommande aux États d'étudier les idées contenues dans le présent rapport et de prendre des mesures concrètes, à la quatre-vingtième session de l'Assemblée générale, pour mettre un place un mécanisme dédié et inclusif permettant d'aborder de manière globale la question de l'IA dans le domaine militaire et ses implications pour la paix et la sécurité internationales**.

Annexe I

Réponses reçues

A. États membres

Allemagne

[Original : anglais]
[11 avril 2025]

I. Introduction

Ces dernières années ont été marquées par une évolution sans précédent des technologies de l'intelligence artificielle (IA), notamment le développement d'applications basées sur des technologies de rupture telles que l'IA générative. Les États doivent impérativement être en mesure de tirer parti des possibilités offertes par ces évolutions technologiques et de veiller à ce que le progrès technologique ne soit pas entravé. Dans le même temps, ils doivent s'assurer que les applications de l'IA dans le domaine militaire seront développées et utilisées de manière responsable et dans le plein respect du droit international, notamment le droit international humanitaire. Les échanges internationaux sont d'une importance capitale pour exceller dans cet exercice d'équilibre.

Dans ce contexte, l'Allemagne contribue activement aux mécanismes internationaux sur les questions liées à l'utilisation responsable de l'IA dans le domaine militaire. L'Allemagne a notamment promu la résolution 79/239 de l'Assemblée générale des Nations Unies sur l'intelligence artificielle dans le domaine militaire et ses conséquences pour la paix et la sécurité internationales, en tant que membre du groupe-noyau des coauteurs de ladite résolution, et soutient pleinement les efforts consentis par le Secrétaire général pour présenter un rapport de fond sur les vues des États Membres concernant « les possibilités et les difficultés que l'application de l'intelligence artificielle dans le domaine militaire présente pour la paix et la sécurité internationales ».

L'Allemagne se réjouit de l'occasion qui lui est donnée d'examiner les vues des États Membres et d'autres parties prenantes et de partager des éléments de sa réflexion sur ces questions importantes.

II. Principes et hypothèses de travail

L'approche adoptée par l'Allemagne pour garantir une utilisation militaire responsable de l'IA repose sur les principes fondamentaux ci-après, définis dans le cadre de divers forums et débats internationaux.

L'Allemagne a activement contribué à l'élaboration des Principes de l'Organisation du Traité de l'Atlantique Nord (OTAN) sur l'utilisation militaire responsable de l'intelligence artificielle en 2021 et reste pleinement alignée sur ces normes importantes : entre autres, la licéité du développement et de l'utilisation d'applications d'intelligence artificielle ; la responsabilité humaine afin de veiller à ce que les personnes qui conçoivent et opèrent l'intelligence artificielle dans les systèmes militaires répondent de leurs actes ; l'explicabilité et la traçabilité des applications de l'intelligence artificielle dans le domaine militaire ; la fiabilité, la sûreté, la sécurité et la robustesse tout au long du cycle de vie des systèmes dotés de l'intelligence et d'autonomie artificielles ; la gouvernabilité, pour assurer l'interaction adéquate entre l'humain et la machine et l'atténuation des biais.

En outre, l'Allemagne a approuvé les documents issus des deux sommets sur l'utilisation responsable de l'IA dans le domaine militaire (REAIM) qui se sont tenus à La Haye en 2023 (Appel à l'action) et à Séoul en 2024 (Plan d'action), ainsi que la Déclaration politique sur l'utilisation responsable de l'intelligence et de l'autonomie artificielles initiée par les États-Unis d'Amérique en 2023, et participe activement au processus de mise en œuvre de la Déclaration.

En outre, l'Allemagne fait également partie de l'initiative de partenariat en matière d'IA au service de la défense, dans le cadre de laquelle des nations partageant les mêmes idées promeuvent l'utilisation responsable de l'IA, font progresser leurs intérêts communs et les meilleures pratiques en matière de mise en œuvre de l'éthique de l'IA, établissent des cadres pour faciliter la coopération et coordonnent les messages stratégiques sur les politiques relatives à l'IA.

En février 2025, l'Allemagne a approuvé la Déclaration de Paris sur le maintien du contrôle humain dans les systèmes d'armes dotés d'IA, qui souligne la nécessité de préserver le contrôle humain dans l'application de l'IA dans le domaine militaire.

III. Aspects clés de l'utilisation de l'intelligence artificielle dans les forces armées fédérales allemandes

Les forces armées fédérales allemandes (Bundeswehr) examinent la possibilité d'utiliser l'IA à la fois pour remplir leur mission principale et pour acquérir une supériorité en matière d'information, de prise de décision et d'efficacité, ainsi que pour optimiser les processus administratifs et logistiques et ceux liés à la maintenance prévisionnelle de systèmes complexes. L'IA est également utilisée pour aider le personnel spécialisé à détecter rapidement les crises civilo-militaires dans le cadre de différentes missions, en analysant des données de masse et en établissant des projections pour les déploiements. L'IA fait partie intégrante des grands projets de défense, qui sont également mis en œuvre dans un contexte européen, contribuant ainsi à maintenir et à promouvoir l'excellence technologique européenne. En ce qui concerne les avancées nationales et technologiques dans le secteur international de l'armement, l'IA sert à faire en sorte que les capacités nécessaires à la défense nationale et alliée soient assurées à l'avenir. Le renforcement des possibilités de déploiement de l'IA, en particulier pour la protection de la sécurité nationale et à des fins militaires, est placé sous l'autorité et la responsabilité des ministères et services compétents. Sans préjudice de ce qui précède, les technologies utilisant l'IA et les applications de l'IA ayant une incidence sur la sécurité sont intégrées dans la stratégie du Gouvernement fédéral allemand en matière d'IA.

La Bundeswehr fixe les exigences éthiques les plus élevées et les normes juridiques les plus strictes en ce qui concerne l'utilisation de l'IA dans les systèmes d'armes. En particulier, la Bundeswehr suit les dispositions du droit international humanitaire en matière de conflits armés et les directives de la Commission d'éthique des données du Gouvernement fédéral et de l'OTAN, notamment les six principes susmentionnés pour une utilisation responsable de l'IA dans le domaine militaire, pendant toute la durée du cycle de vie de ces systèmes.

IV. Considérations essentielles

Afin de maintenir les capacités de défense et de dissuasion nécessaires, l'Allemagne reste déterminée à saisir les possibilités offertes par l'IA dans le domaine militaire et est convaincue que le progrès technologique ne doit pas être entravé, en particulier compte tenu de la double utilisation inhérente aux technologies en question.

Parallèlement, l'Allemagne continuera d'élargir sa base de connaissances en évaluant et en traitant les risques associés à l'utilisation de l'IA dans le domaine

militaire, notamment ceux liés aux biais involontaires, tels que ceux fondés sur le genre. Dans ce contexte, l'Allemagne attache une grande importance au rôle essentiel du monde universitaire et aux contributions précieuses des instituts de recherche et des groupes de réflexion travaillant dans ce domaine. Afin d'encourager les travaux de recherche, l'Allemagne soutient les organisations de recherche concernées, notamment l'Institut des Nations Unies pour la recherche sur le désarmement (UNIDIR), en contribuant financièrement à des projets de recherche appliquée.

Il est primordial pour l'Allemagne de garantir l'inclusivité des discussions, tant en veillant à l'équilibre géographique qu'en tenant compte des points de vue des États Membres, mais aussi du secteur, de la société civile et du monde universitaire.

Dans le cadre de l'examen des perspectives et des risques liés aux systèmes d'armes basés sur l'IA, l'Allemagne attache une importance particulière au concept de contrôle humain et considère l'existence d'un cadre efficace de contrôle humain comme une condition nécessaire pour garantir que tous les systèmes d'armes sont conformes au droit international humanitaire. Pour ce faire, il convient non seulement d'exercer un contrôle technique, mais aussi de faire preuve de discernement. Le concept allemand d'accord-cadre pour la maîtrise technologique englobe un ensemble de mesures et d'actions réalisables sur le plan technologique qui fixent des limites claires à l'intérieur desquelles l'algorithme du système peut fonctionner. Le droit international, et en particulier le droit international humanitaire, est un élément central de ces limites. Lorsqu'il s'agit de l'utilisation réelle de l'IA sur le champ de bataille, le contexte est de la plus haute importance. L'Allemagne estime que le concept de cadre de contrôle humain est un moyen approprié de tenir compte de cet aspect de manière adéquate.

Une attention particulière est nécessaire lorsque l'utilisation de l'IA est liée aux armes nucléaires, un domaine dans lequel le débat scientifique et politique n'en est qu'à ses prémices. L'utilisation éventuelle de l'IA dans les systèmes de commande et de contrôle des armes nucléaires pourrait avoir de graves répercussions sur la stabilité stratégique ou l'escalade nucléaire. Dans le même temps, l'IA pourrait ouvrir de nouvelles perspectives en matière de lutte contre la prolifération et l'utilisation des armes de destruction massive. L'Allemagne a cherché à contribuer à ces débats en accueillant une conférence sur l'intelligence artificielle et les armes de destruction massive dans le cadre de sa série de conférences bien connue consacrée au thème « Capturing technology – rethinking arms control » (Tirer parti de la technologie – repenser la maîtrise des armements), qui s'est tenue à Berlin le 28 juin 2024.

La Convention sur l'interdiction de la mise au point, de la fabrication et du stockage des armes bactériologiques (biologiques) ou à toxines et sur leur destruction et la Convention sur l'interdiction de la mise au point, de la fabrication, du stockage et de l'emploi des armes chimiques et sur leur destruction interdisent expressément l'emploi de telles armes. Les applications telles que les grands modèles de langage (génératifs) peuvent faciliter la prolifération de connaissances à double usage susceptibles d'être utilisées à mauvais escient pour développer, produire ou utiliser des armes biologiques et chimiques. La convergence des applications de l'IA, telles qu'AlphaFold, et de la biologie de synthèse peut permettre à des acteurs malveillants de concevoir de nouvelles protéines qui, grâce à des modifications de la séquence d'ADN, peuvent échapper à la détection. L'IA peut être utilisée pour analyser des mégadonnées stockées dans le cloud, telles que les données sur le génome humain, et présenter de grands avantages pour le développement de thérapies médicales individuelles, mais elle pourrait également être utilisée à mauvais escient pour développer des armes biologiques ciblant des groupes ethniques précis.

En étroite collaboration avec ses partenaires internationaux, l'Allemagne continuera donc de rechercher les lignes d'action possibles pour évaluer l'incidence

des applications de l'IA sur le développement et la production d'armes interdites et lancer d'éventuelles réglementations. Dans le même temps, l'Allemagne exploitera les avantages de l'IA pour la vérification, la bioinformatique et la réduction des risques.

V. L'engagement de l'Allemagne dans les processus internationaux

Depuis le début, l'Allemagne a contribué activement au processus relatif à l'utilisation responsable de l'intelligence artificielle dans le domaine militaire et continuera de le faire. L'Allemagne a fait partie du groupe-noyau des coauteurs de la résolution 79/239 de l'Assemblée générale relative à l'intelligence artificielle dans le domaine militaire et ses conséquences pour la paix et la sécurité internationales. L'Allemagne se félicite de l'approche interrégionale et multipartite adoptée dans le cadre de cette initiative majeure et attend avec impatience qu'elle se poursuive en Espagne en septembre 2025.

À titre de complément, l'Allemagne a contribué à la Déclaration politique sur l'utilisation responsable de l'intelligence et de l'autonomie artificielles, initiée par les États-Unis, notamment en coprésidant le groupe de travail chargé de la supervision (conjointement avec l'Autriche).

En outre, l'Allemagne est activement engagée dans l'initiative de partenariat en matière d'IA au service de la défense et participe au réseau d'experts de l'UNIDIR sur la gouvernance de l'intelligence artificielle dans le domaine militaire.

L'Allemagne soutient l'Ambassadeur Robert in den Bosch, Président du Groupe d'experts gouvernementaux sur les technologies émergentes dans le domaine des systèmes d'armes létaux autonomes à Genève, et reste activement engagée dans le processus, notamment en coordonnant les positions de plusieurs États Membres dans le cadre de ce que l'on appelle l'approche à deux niveaux. En étroite collaboration avec ses partenaires internationaux, l'Allemagne continuera d'œuvrer à l'accomplissement du mandat du Groupe dans les délais impartis, de préférence d'ici à la fin de 2025.

Dans le contexte de l'OTAN, l'Allemagne reconnaît le potentiel de l'IA au service du renforcement des forces armées et des capacités de défense de l'Alliance, ainsi que les difficultés que l'utilisation de l'IA posera en ce qui concerne l'interopérabilité des forces armées des pays membres de l'Alliance. Les évolutions multinationales de l'IA et les aspects de normalisation de l'IA au sein de l'OTAN, de l'Union européenne et des pays partenaires de l'Allemagne doivent être pleinement pris en compte, afin de garantir l'interopérabilité de la Bundeswehr en tant que force militaire dans le cadre d'opérations internationales. Par conséquent, l'Allemagne s'est félicitée que les pays de l'OTAN se soient mis d'accord sur les principes d'utilisation responsable dans le cadre de la stratégie de l'OTAN en matière d'IA.

VI. Voie à suivre

Étant donné que les nouvelles technologies de rupture continueront d'évoluer et de façonner notre monde, l'Allemagne considère qu'une coordination internationale inclusive sur le développement et l'utilisation militaires responsables de l'IA est indispensable. Les processus internationaux existants fournissent un excellent cadre pour traiter les aspects significatifs en jeu et pour prendre en compte les points de vue d'une variété de parties prenantes concernées. L'Allemagne continuera de contribuer activement à ces efforts afin de mettre en œuvre les engagements politiques en matière d'utilisation responsable de l'IA dans le domaine militaire, tels que la Déclaration politique initiée par les États-Unis et le processus relatif à l'utilisation responsable de l'IA dans le domaine militaire, et d'élargir le soutien qui leur est apporté. L'Allemagne attend avec intérêt de prendre connaissance des conclusions du rapport

du Secrétaire général sur l'IA dans le domaine militaire. Elle continuera de contribuer activement au processus relatif aux systèmes d'armes létaux autonomes dans le cadre du Groupe d'experts gouvernementaux à Genève.

Argentine

[Original : espagnol]
[10 avril 2025]

Comme suite à la résolution 79/239 intitulée « L'intelligence artificielle dans le domaine militaire et ses conséquences pour la paix et la sécurité internationales », adoptée par l'Assemblée générale le 24 décembre 2024, les vues de l'Argentine sont exposées ci-après.

Approche globale

La République argentine constate que l'évolution de l'intelligence artificielle (IA) dans le domaine militaire a des retombées stratégiques notables. Son utilisation comporte des avantages concrets dans un certain nombre de fonctions non létales, tout en induisant des risques qui requièrent une prise en compte au niveau du droit international, de l'éthique et de la responsabilité opérationnelle. Dans ce cadre, le développement et l'utilisation de ces technologies doivent respecter le droit international humanitaire et les droits humains, en garantissant à tout moment la responsabilité et le contrôle exercés par l'être humain dans la prise de décisions critiques.

Possibilités

L'IA de défense, en particulier dans ses applications non létales, est un outil légitime et précieux pour l'amélioration des capacités nationales. Parmi les utilisations prioritaires, on peut citer les suivantes :

- optimisation de la logistique et des opérations,
- appui au traitement du renseignement,
- renforcement de la cyberdéfense,
- simulation, formation et planification stratégique.

Ces capacités contribuent à rendre les opérations plus efficaces, plus sûres et mieux adaptées aux scénarios contemporains, en améliorant l'efficacité défensive sans compromettre les principes humanitaires ou les obligations internationales de l'État.

Difficultés

Le développement accéléré de l'IA dans le contexte militaire crée des défis qui doivent être relevés collectivement, tels que :

- l'abaissement du seuil de recours à la force et la diminution du temps de décision par les humains,
- l'existence potentielle de biais algorithmiques non détectés,
- la prolifération de systèmes autonomes au profit d'agents non étatiques,
- le risque de renforcement des écarts technologiques entre les États,
- Ces risques rendent d'autant plus nécessaire l'adoption de principes communs, de garanties vérifiables et de cadres de coopération.

Gouvernance, coopération internationale et inclusion technologique

Nous pensons que tout processus réglementaire dans ce domaine doit reposer sur les principes suivants :

- éviter les réglementations générales ou prématurées qui limitent le développement autonome de technologies de défense légitimes,
- faire une distinction claire entre les utilisations létales et non létales,
- garantir un contrôle humain marqué, en tant qu'exigence opérationnelle et réglementaire essentielle,
- promouvoir une coopération internationale inclusive visant à renforcer les capacités et à réduire les écarts technologiques entre les États.

L'Argentine a réaffirmé ces principes lors de récents forums multilatéraux, en insistant sur l'importance de promouvoir des normes communes pour l'utilisation responsable de l'IA dans le domaine militaire, en particulier en ce qui concerne la cyberdéfense et la cybersécurité.

À titre d'exemple d'initiative menée au niveau régional, la réunion du comité de travail sur le développement, l'application et la gouvernance responsables de l'intelligence artificielle dans le domaine militaire s'est tenue à Mendoza (en Argentine), dans le cadre de la XVI^{e} Conférence des Ministres de la défense des Amériques en 2024, dans le but d'œuvrer de concert à l'élaboration de normes internationales.

Référence au Pacte pour l'avenir

Enfin, la République argentine s'est formellement désolidarisée du Pacte pour l'avenir, mentionné dans le préambule de la résolution 79/239 de l'Assemblée générale. Cette mention ne représente donc pas un engagement, une adhésion ou une approbation de la part de l'État argentin.

Autriche

[Original : anglais]
[11 avril 2025]

Conformément à la demande formulée au paragraphe 7 de la résolution 79/239 de l'Assemblée générale, l'Autriche souhaite partager les réflexions et observations recueillies à l'échelle du pays, comme suit :

L'intelligence artificielle appliquée à la cybersécurité et à la cyberdéfense

Les logiciels de cybersécurité utilisant l'intelligence artificielle (IA) sont déjà largement utilisés pour aider à détecter les intrusions et autres activités malveillantes dans les réseaux informatiques. Ces outils d'IA permettront probablement d'automatiser davantage la protection des systèmes informatiques, en recherchant les vulnérabilités et les activités suspectes afin d'améliorer la résilience des logiciels et du matériel.

Dans le même temps, les outils d'IA sont de plus en plus utilisés pour accroître le degré de sophistication des cyberattaques et créer de nouveaux virus informatiques, dans une course entre modèles défensifs et offensifs basés sur l'IA en matière de cybersécurité. En outre, les logiciels utilisant l'IA, notamment les grands modèles de langage, facilitent l'accès aux acteurs mal intentionnés, qui peuvent de plus en plus créer des logiciels malveillants sans avoir besoin de compétences approfondies en programmation.

L’intelligence artificielle appliquée aux campagnes de désinformation en tant qu’élément de stratégies hybrides

Les logiciels utilisant l’IA pour créer et diffuser des contenus falsifiés sont de plus en plus utilisés pour renforcer les campagnes de désinformation. L’IA générative est notamment utilisée pour diffuser à grande échelle du contenu personnalisé et adapté aux réalités locales. En outre, les logiciels utilisant l’IA et destinés à produire des hypertrucages (*deepfakes*) audio ou vidéo s’améliorent rapidement et sont déjà largement utilisés. Les contenus falsifiés ainsi produits peuvent être diffusés à l’aide de réseaux massifs de robots pilotés par l’IA dans les médias sociaux, afin de donner le sentiment que l’opinion publique bascule. L’IA réduit donc les obstacles à la conduite de campagnes de désinformation à grande échelle, car la quantité et la qualité des faux contenus créés ne sont plus limitées par le nombre ou les compétences d’opérateurs humains.

Toutefois, les algorithmes d’IA peuvent également être utilisés pour détecter les contenus générés par l’IA et les campagnes de similitantisme, tandis que les fichiers audio et vidéo de type *deepfake*, qui semblent parfaitement authentiques, peuvent être mis en évidence à l’aide d’outils d’IA spécialisés. Il est nécessaire d’utiliser ces outils pilotés par l’IA pour contrer les effets néfastes de l’IA utilisée à des fins de campagnes de désinformation.

L’intelligence artificielle appliquée à la prolifération des armes

L’IA peut réduire les obstacles à l’acquisition d’armes, y compris d’armes de destruction massive. En raison de leur capacité à fournir des compétences spécialisées sur simple pression d’un bouton, les grands modèles de langage et les applications qui en découlent pourraient faciliter la fabrication d’armes par des acteurs malveillants. En effet, ils peuvent notamment permettre d’accéder à des plans de fabrication ou d’impression de composantes d’armes légères et de petit calibre ou encore de modifier des agents pathogènes à des fins de guerre biologique. Si la facilité d’accès aux connaissances réduit la portée et l’ampleur des programmes d’armement, il sera plus difficile de détecter ces menaces, de les prévenir et de s’y préparer.

Dans le même temps, les algorithmes d’apprentissage automatique peuvent également être utilisés pour lutter contre la prolifération des armes. Grâce à leurs capacités de détection des anomalies et de reconnaissance des formes, ils peuvent aider à mettre au jour les activités malveillantes, notamment en détectant les flux d’argent illicites destinés aux programmes d’armement ou en analysant les formes des données satellitaires.

L’intelligence artificielle appliquée à la vérification de la maîtrise des armements et à la prise de décision en cas de situations de crise

Du fait de sa capacité à analyser de grandes quantités de données, provenant par exemple d’images satellites, et à classer différents objets, l’IA peut contribuer à la vérification des accords de maîtrise des armements. Elle permet de repérer des équipements militaires tels que des chars, des missiles et des casernes, ou encore des activités militaires telles que des mouvements de contingents ou des exercices militaires. En outre, comme nous l’avons déjà mentionné, les programmes d’armement illégaux pourraient être détectés plus facilement grâce à l’IA. Il serait donc beaucoup plus difficile d’enfreindre les accords sur la maîtrise des armements et les États parties pourraient être sûrs que tout le monde en respecte les dispositions des accords.

Des informations plus nombreuses et de meilleure qualité, basées sur la capacité de l’IA à analyser et à classer les données issues des capteurs, peuvent non seulement faciliter la mise en œuvre des accords relatifs à la maîtrise des armements, mais

également contribuer à une meilleure prise de décision dans les situations où les tensions militaires entre États sont particulièrement vives. Grâce à l'IA, les dirigeants politiques et militaires pourraient bénéficier d'une meilleure perception de la situation et ainsi désamorcer les crises.

La paix et la sécurité et la Charte des Nations Unies

Le recours à l'IA dans le domaine militaire soulève un problème particulier, à savoir les risques potentiels pour la paix et la sécurité d'une escalade involontaire ou de malentendus résultant de l'utilisation de cette technologie. Le recours à l'apprentissage automatique ajoute un degré de complexité supplémentaire, car il se peut que certaines parties prenantes ne comprennent pas entièrement le fonctionnement d'un système.

En ce qui concerne l'interaction humain-machine et la nécessité d'une intervention humaine, il est également nécessaire de mettre en place des mesures et des garde-fous pour garantir l'application du principe de responsabilité et atténuer les biais algorithmiques dans le cadre de l'utilisation de l'IA dans les systèmes informatisés d'aide à la décision.

Tous ces risques doivent être atténués par une surveillance et des mesures qui tiennent compte des difficultés inhérentes à ces technologies.

Il convient de noter que l'Article 36 du Protocole additionnel aux Conventions de Genève du 12 août 1949 relatif à la protection des victimes des conflits armés internationaux (Protocole I) impose l'obligation d'examiner la légalité d'une nouvelle arme, de nouveaux moyens ou d'une nouvelle méthode de guerre avant qu'ils soient utilisés dans un conflit armé.

L'IA peut également être utilisée pour favoriser la mise en œuvre effective des obligations découlant du droit international humanitaire, en particulier en ce qui concerne la protection des civils, en tant qu'obligation positive et action positive, notamment dans le cadre de projets, de travaux de recherche et d'applications expressément conçues à cet effet.

Cadres relatifs à la coopération multilatérale et à l'échange d'informations

La question de l'IA dans le domaine militaire évolue rapidement et soulève des difficultés pour tous les États. Les discussions multilatérales et les formats d'échange d'expériences et de bonnes pratiques sont donc tout à fait pertinents. À cet égard, l'Autriche a approuvé la Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles. En tant que coprésidentes du Groupe de travail chargé de superviser cette Déclaration, l'Autriche et l'Allemagne ont facilité le partage des meilleures pratiques destinées à résoudre les difficultés et à formuler des politiques dans ce domaine. L'Autriche a également approuvé le plan d'action issu du Sommet sur l'intelligence artificielle responsable dans le domaine militaire (REAIM) ainsi que la Déclaration de Paris sur le maintien du contrôle humain dans les systèmes d'armes dotés d'IA.

Relation entre les travaux de la communauté internationale relatifs à l'intelligence artificielle dans le domaine militaire et ses travaux relatifs aux systèmes d'armes autonomes

Dans le cadre plus large de l'application de l'intelligence et de l'autonomie artificielles dans le domaine militaire, il convient de souligner la question des systèmes d'armes autonomes, qui soulèvent des préoccupations particulières d'un point de vue juridique, éthique et de sécurité. Cette question n'est pas au cœur de la résolution 79/239 de l'Assemblée générale des Nations Unies, car les discussions

menées dans le cadre de l'ONU se poursuivent déjà depuis 2013, une majorité croissante d'États ayant exprimé leur souhait d'établir des règles et des limites aux systèmes d'armes autonomes au niveau international. Dans le présent rapport, l'Autriche se limitera donc à préciser sa position en faveur d'un instrument juridiquement contraignant sur les systèmes d'armes autonomes et à faire référence aux travaux importants entrepris actuellement par le Groupe d'experts gouvernementaux dans le cadre de la Convention sur l'interdiction ou la limitation de l'emploi de certaines armes classiques qui peuvent être considérées comme produisant des effets traumatiques excessifs ou comme frappant sans discrimination, ainsi qu'aux efforts complémentaires déployés dans le cadre de la toute première résolution de l'Assemblée générale des Nations Unies sur les systèmes d'armes létaux autonomes (résolution 78/241), à la suite de laquelle le Secrétaire général de l'Organisation des Nations Unies a publié un rapport (A/79/88) et de sa résolution de suivi (résolution 79/62), qui prévoit que des consultations informelles sur les systèmes d'armes létaux autonomes se tiendront à New York les 12 et 13 mai 2025.

Considérations relatives aux cadres juridiques régissant l'intelligence artificielle

Le Règlement sur l'intelligence artificielle de l'Union européenne établit un cadre législatif pour les systèmes d'IA dans différents secteurs et vise à renforcer la confiance à l'égard des applications de l'IA et à exploiter les avantages de cette technologie tout en sauvegardant les droits humains, les libertés fondamentales et les valeurs démocratiques. Il souligne l'importance de la transparence, de l'application du principe de responsabilité et de la surveillance humaine dans le développement et le déploiement des systèmes d'IA, tout en promouvant la sécurité juridique, l'innovation et la compétitivité. Il ne s'applique pas aux systèmes d'IA développés pour des activités militaires, de défense ou de sécurité nationale. Il repose toutefois sur une approche fondée sur les risques, qui pourrait être utile pour gérer le large éventail d'applications potentielles de l'IA dans le domaine militaire.

Voie à suivre

L'Autriche salue les travaux entrepris dans les différents formats et forums mentionnés dans sa contribution concernant les applications de l'IA dans le domaine militaire et est convaincue qu'ils contribueront à l'émergence d'un ensemble de normes internationalement reconnues visant à garantir une utilisation responsable de l'IA dans ce domaine, conformément aux obligations juridiques internationales et aux principes éthiques.

Chili

[Original : espagnol]
[11 avril 2025]

Le Chili a déjà souligné que le développement rapide de technologies nouvelles et émergentes est un aspect essentiel de la sécurité internationale et représente un défi pour tous les pays. Ces nouvelles technologies, et en particulier l'intelligence artificielle (IA), peuvent s'avérer très bénéfiques pour le développement et le bien-être des sociétés, mais soulèvent en même temps d'importantes questions sur les incidences de leur utilisation dans les domaines de la sécurité et de la défense. L'utilisation de ces nouvelles technologies peut comporter des avantages de taille, mais aussi des risques et des difficultés.

À cet égard, le Chili estime qu'il est souhaitable de parvenir à une compréhension commune de l'utilisation responsable de l'IA dans le domaine militaire et de la sécurité, ainsi que du développement et de l'utilisation de systèmes d'armes létaux autonomes. Il soutient les efforts multilatéraux visant à créer et à

renforcer les instances de dialogue et de discussion entre les pays afin de parvenir à un consensus et à une compréhension mutuelle pour ce qui concerne l'utilisation de ces nouvelles technologies.

Le Chili a acquis une position de leader dans le domaine de l'intelligence artificielle, tant en raison de ses progrès notables dans les conditions propices au déploiement de la technologie que de son rôle de pionnier en matière de politique d'intelligence artificielle et de débat sur la réglementation. En octobre 2021, le Chili a lancé sa première politique nationale en matière d'IA, élaborée en collaboration avec divers acteurs publics et privés. Cette politique est axée sur trois domaines essentiels : les facteurs favorables, l'utilisation et le développement de la technologie, et la mise en place de cadres réglementaires et éthiques visant à garantir l'utilisation responsable et sûre de l'IA.

En 2024, le Chili a lancé une version actualisée de la Politique nationale en matière d'IA, qui intègre de nouveaux sous-axes tels que l'articulation internationale, l'environnement et la crise climatique, l'inclusion et la non-discrimination, les enfants et les adolescents, ainsi que la culture et la préservation du patrimoine culturel. La Politique est accompagnée du plan d'action correspondant, qui comprend plus de 100 mesures engagées pour l'année 2026 dans des domaines tels que l'éducation, la santé, l'environnement et la culture. La nouvelle Politique intègre aussi les principes de la Recommandation sur l'éthique de l'intelligence artificielle formulée par l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO), de respecter les cadres internationaux les plus récents.

Il faut noter que le Chili a été le premier pays au monde à achever la méthode d'évaluation de l'état de préparation, instrument développé par l'UNESCO pour mesurer l'état de préparation d'un pays pour ce qui est du déploiement éthique et responsable de l'IA. Il a ainsi réaffirmé sa volonté d'appliquer la Recommandation sur l'éthique de l'intelligence artificielle de l'UNESCO dans l'environnement politique national. Le pays a encouragé le développement éthique et responsable de la technologie, ce qui s'est traduit par sa participation aux sommets sur l'IA organisés par le Royaume-Uni (2023), la République de Corée (2024) et la France (2025).

Au niveau législatif, le Chili examine actuellement un projet de loi visant à réglementer les systèmes d'IA, sur la base d'une approche fondée sur les risques, afin que ceux-ci puissent être développés et déployés dans le respect des principes démocratiques et des droits fondamentaux des personnes.

Dans le domaine de la défense et de la sécurité, le Chili a soutenu les Sommets sur l'intelligence artificielle responsable dans le domaine militaire, organisés à La Haye (2023) et à Séoul (2024), et y a activement participé. Le pays a souscrit aux documents adoptés lors des deux sommets [« Call to action » (2023) et « Blueprint for action » (2024)]. Il soutient aussi les travaux menés par la Commission mondiale sur l'intelligence artificielle responsable dans le domaine militaire (Global Commission on Responsible Artificial Intelligence in the Military Domain).

Il importe aussi de noter qu'un atelier régional sur l'utilisation responsable de l'intelligence artificielle dans le domaine militaire et dans celui de la sécurité en général, organisé par les Ministères des affaires étrangères du Chili et du Costa Rica, sous l'égide du Royaume des Pays-Bas et de la République de Corée, s'est tenu au Chili les 13 et 14 juin 2024. La manifestation a aussi reçu l'appui du Centro de Estudios en Derecho, Tecnología y Sociedad de l'Université du Chili et du Centre pour le dialogue humanitaire, basé à Genève. Des représentants de l'Argentine, du Brésil, du Chili, de la Colombie, du Costa Rica, de l'Équateur, de la Jamaïque, du Mexique, du Paraguay, de la République dominicaine, du Salvador, de la Trinité-et-Tobago et de l'Uruguay, ainsi que du Royaume des Pays-Bas et de la République de

Corée ont participé à l'atelier. Des représentants du Ministère chilien de la défense nationale et des forces armées étaient également présents.

Le Chili estime que l'IA appliquée au domaine militaire et à celui de la sécurité peut créer des possibilités et des avantages, tels que l'amélioration de la prise de décision et de l'analyse stratégique, la réalisation d'opérations logistiques plus efficaces, l'amélioration des capacités de cyberdéfense et de cybersécurité par le renforcement de la sécurité des infrastructures critiques, ainsi que l'aide à la planification de missions complexes de maintien de la paix et d'aide humanitaire, ou l'amélioration des capacités de vérification et de contrôle dans le domaine de la maîtrise des armements et du respect des régimes de maîtrise des armements.

Le Chili estime que les technologies liées à l'IA doivent être développées, déployées et utilisées dans le respect du droit international, y compris, le cas échéant, de la Charte des Nations Unies, du droit international humanitaire, du droit international des droits humains et d'autres cadres juridiques applicables.

Pour le Chili, des mesures de contrôle et de sécurité doivent être mises en place pour empêcher des acteurs irresponsables d'acquérir et d'utiliser à mauvais escient des capacités d'IA potentiellement nuisibles dans le domaine militaire, y compris des systèmes fondés sur l'IA. Ces mesures ne doivent cependant pas compromettre l'accès équitable aux avantages créés par les capacités de l'IA dans d'autres domaines non militaires.

De même, le Chili estime qu'il est indispensable d'unir les efforts pour empêcher que les technologies de l'IA ne soient utilisées pour contribuer à la prolifération des armes de destruction massive par des acteurs étatiques et non étatiques, y compris des groupes terroristes, et d'insister sur le fait que les technologies de l'IA devraient soutenir et non entraver les efforts menés en faveur du désarmement, de la maîtrise des armements et de la non-prolifération. Il importe particulièrement de maintenir le contrôle humain et la participation à toutes les activités permettant d'éclairer et d'appliquer les décisions souveraines relatives à l'utilisation des armes nucléaires, sans préjudice de l'objectif final d'un monde exempt d'armes nucléaires.

Le Chili encourage l'élaboration de mesures de confiance, telles que l'échange d'informations et les consultations portant sur les bonnes pratiques et les enseignements tirés entre les États. À cet égard, le Chili estime que les pays doivent pouvoir élaborer et mettre en place des stratégies, des principes, des normes, des politiques, des cadres et des textes législatifs au niveau national afin que l'IA soit utilisée de façon responsable dans le domaine militaire. Les mesures de confiance peuvent constituer un outil efficace pour développer des mécanismes d'endiguement, de contrôle et de crédibilité, tant au niveau national qu'international, en faisant la promotion de la transparence.

De même, le Chili considère qu'il faut réduire la fracture numérique et la fracture relative à l'IA qui existent entre les pays développés et les pays en développement, et améliorer la compréhension et la sensibilisation pour ce qui concerne les incidences de l'intelligence artificielle dans le domaine militaire, sans oublier l'échange de connaissances et le partage des bonnes pratiques et des enseignements tirés entre tous les États.

À cet égard, le Chili estime qu'il est indispensable de pouvoir développer des initiatives et des programmes qui favorisent le renforcement des capacités, en particulier dans les pays en développement, afin de promouvoir la pleine participation de ces pays aux discussions sur la gouvernance de l'IA militaire, en reconnaissant que le renforcement des capacités peut aussi aider les pays à approfondir leur compréhension de l'IA militaire et à faciliter le développement, le déploiement et

l'utilisation de telles capacités d'une manière responsable et légale. Le renforcement des capacités permettra également aux pays de mieux aborder les discussions et dialogues internationaux.

Le Chili considère qu'il importe de consolider la coopération internationale en matière de renforcement des capacités, en encourageant le dialogue et le débat aux niveaux national, régional, sous-régional et interrégional, y compris les programmes de formation, les conférences, les ateliers et les séminaires, entre autres, pour les fonctionnaires aux niveaux diplomatique, politique et technique, afin de réduire le déficit de connaissances en matière de développement, de déploiement et d'utilisation responsables de l'intelligence artificielle dans le domaine militaire.

Le Chili est conscient de l'intérêt des débats et dialogues régionaux et sous-régionaux sur l'IA dans le domaine militaire et considère qu'il est primordial de les promouvoir. À cet égard, il cite la XVI^{e} Conférence des Ministres de la défense des Amériques, tenue en Argentine du 13 au 16 octobre 2024, et en particulier la Déclaration de Mendoza, document publié à l'issue de cette Conférence, qui recommande, entre autres, de promouvoir l'utilisation éthique de l'IA dans le domaine de la défense, de tenir compte de la diversité économique et technologique des pays membres de la Conférence, et de promouvoir des mécanismes de renforcement de la confiance mutuelle et de la coopération hémisphérique et régionale qui permettent aux pays membres de la Conférence de partager leurs connaissances et leurs meilleures pratiques et d'élaborer des normes faisant consensus, ainsi que de développer des capacités technologiques aux fins de l'utilisation de l'IA dans le domaine de la défense.

Enfin, le Chili estime qu'il est essentiel que toutes les parties prenantes, telles que la société civile, les universités, l'industrie, le secteur privé, les experts techniques et les organisations régionales et internationales, participent à la discussion et au dialogue sur l'utilisation de l'IA dans le domaine militaire.

Chine

[Original : chinois]
[11 avril 2025]

Le développement rapide et l'application généralisée de l'intelligence artificielle (IA) dans le domaine militaire remodèlent les paradigmes de la guerre de demain tout en créant des risques pour la paix et la sécurité internationales. Alors que le monde doit faire face à de multiples défis lancés à la paix et la sécurité, toutes les parties devraient rechercher un consensus, par le dialogue et la coopération, sur la réglementation des applications militaires de l'IA, promouvoir l'élaboration d'un dispositif de gouvernance ouvert, équitable et efficace pour la sécurité de l'IA, et minimiser les risques afin de garantir que les technologies utilisant l'IA restent sûres, fiables et contrôlables et qu'elles se développent toujours d'une manière qui profite au progrès de la civilisation humaine.

La Chine s'est toujours engagée de manière responsable et constructive dans la gouvernance mondiale des applications militaires de l'IA. Nous préconisons l'adoption d'une approche axée sur l'être humain dans les applications militaires de l'IA et la promotion de la vision d'une sécurité commune, globale, coopérative et durable, afin de bâtir une communauté d'avenir partagé pour l'humanité. En 2021, dans le cadre de la Convention sur certaines armes classiques, la Chine a présenté un document de position sur la réglementation des applications militaires de l'IA, qui proposait des points de vue et des recommandations systématiques sur le développement et l'utilisation responsables de l'IA dans le domaine militaire en matière de sécurité stratégique, de politiques militaires, de droit et d'éthique, de

sécurité technologique, d'opérations de recherche et de développement, de gestion et de contrôle des risques, d'élaboration de règles et de coopération internationale. En 2023, la Chine a proposé l'Initiative mondiale pour la gouvernance de l'intelligence artificielle, dans laquelle tous les pays, en particulier les grandes puissances, étaient invités à adopter une attitude prudente et responsable à l'égard de la recherche, du développement et de l'application des technologies de l'IA dans le domaine militaire. Les propositions concrètes qui ont été formulées sont notamment les suivantes.

Premièrement, il convient d'adopter une approche prudente et responsable. Tout en développant leurs capacités de défense nationales légitimes, tous les pays, en particulier les grandes puissances, devraient s'abstenir de rechercher une supériorité militaire absolue au moyen de l'IA et de porter atteinte aux intérêts légitimes des autres en matière de sécurité. Il faut s'efforcer d'éviter les malentendus et les erreurs de calcul et de prévenir une course aux armements dans ce domaine.

Deuxièmement, il faut adopter une approche axée sur l'être humain. Il est essentiel de toujours considérer les êtres humains comme le sujet responsable ultime et de veiller à ce que les systèmes d'armes concernés soient placés sous contrôle humain. Les applications militaires de l'IA doivent respecter et protéger la dignité humaine et les droits humains et honorer les valeurs communes de l'humanité.

Troisièmement, il convient de respecter le principe de base « AI for good » (l'IA au service du bien). L'application de l'IA dans le domaine militaire doit contribuer au maintien de la paix, respecter le droit international humanitaire et les autres lois internationales applicables, et viser à réduire les pertes collatérales.

Quatrièmement, il convient de mettre en place une gouvernance agile. Nous devrions renforcer l'évaluation prospective des risques et la formation du personnel à l'IA, prendre les mesures d'atténuation des risques nécessaires et réduire les risques de prolifération, sans pour autant entraver l'innovation et l'utilisation pacifique des technologies.

Cinquièmement, le multilatéralisme doit être maintenu. Nous devons aider l'Organisation des Nations Unies à remplir le rôle qui lui incombe, saluer la mise en place de plateformes de discussion ouvertes à toutes les parties et nous efforcer de créer des dispositifs de gouvernance fondés sur une participation universelle et un large consensus.

La Chine estime que l'importance de l'IA dans le domaine militaire doit être évaluée objectivement. Il est essentiel d'orienter le développement de l'IA militaire dans une direction appropriée tout en empêchant une croissance non réglementée. Dans la prochaine phase, la communauté internationale devrait collaborer pour maximiser les avantages tout en minimisant les inconvénients. La Chine propose les idées et suggestions décrites ci-après.

Tout d'abord, il faut établir des lignes directrices claires. La sécurité et le développement doivent faire l'objet d'une attention égale. Il est impératif de respecter les buts et principes énoncés dans la Charte des Nations Unies ainsi que les normes fondamentales régissant les relations internationales, et de veiller à ce que la technologie de l'IA ne devienne pas un outil utilisé pour envahir les autres pays et servir des visées hégémoniques. La Chine est disposée à poursuivre avec toutes les parties les échanges portant sur une approche axée sur l'être humain dans les applications militaires de l'IA et à s'employer sans relâche à dégager un consensus.

Ensuite, il est nécessaire d'améliorer les mesures de gouvernance. Dans le contexte actuel de développement et d'application de l'IA, nous devrions promouvoir la mise en place d'un système de test et d'évaluation, mettre en œuvre une gouvernance agile et opter pour une gestion par niveaux et par catégories pour une réponse rapide et efficace. Tous les pays devraient, en fonction de leur situation

nationale, établir et améliorer les systèmes juridiques et réglementaires nationaux, affiner les lignes directrices éthiques applicables et renforcer l'éducation et la formation, afin d'améliorer la sécurité, la fiabilité et la contrôlabilité des technologies utilisant l'IA.

Enfin, il faut renforcer la coopération internationale. Tous les pays devraient adhérer aux principes d'ouverture et d'inclusion, participer aux dialogues et aux échanges pour améliorer la compréhension mutuelle, et renforcer la coordination des politiques et la coopération en matière de renforcement des capacités pour ce qui concerne la gouvernance de l'IA afin d'améliorer continuellement le niveau de gouvernance.

Égypte

[Original : anglais]
[11 avril 2025]

Conformément à la résolution 79/239 de l'Assemblée générale, le Gouvernement de la République arabe d'Égypte souhaite faire part de son point de vue sur les possibilités et les difficultés que l'application de l'intelligence artificielle dans le domaine militaire présente pour la paix et la sécurité internationales.

La résolution 79/239 de l'Assemblée générale représente une étape importante dans la promotion du multilatéralisme sur le thème de l'IA dans le domaine militaire et vise à inscrire cette question en tête de l'ordre du jour politique. Elle fait suite à l'appel lancé par le Secrétaire général de l'Organisation des Nations Unies en faveur de l'élaboration de normes, de règles et de principes relatifs à la conception, au développement et à l'utilisation d'applications militaires de l'IA, avec la participation de toutes les parties prenantes concernées.

Étant entendu que la résolution susmentionnée, en vertu de laquelle ces points de vue sont présentés, vise à mettre l'accent sur des domaines autres que les systèmes d'armes létaux autonomes, il est essentiel de rappeler la position inébranlable de l'Égypte selon laquelle toute discussion sérieuse sur le sujet ne doit en aucun cas faire l'impasse sur la priorité d'aborder toutes les dimensions éthiques, juridiques et de sécurité entourant les systèmes d'armes létaux autonomes, qui représentent la menace la plus pressante pour le maintien de la paix et de la sécurité internationales eu égard aux applications militaires de l'intelligence artificielle.

L'adoption d'une interdiction juridiquement contraignante des systèmes d'armes létaux autonomes fonctionnant sans contrôle ni surveillance humaine et ne pouvant être utilisés conformément au droit international humanitaire, comme l'a suggéré le Secrétaire général, est la solution la plus efficace et la plus réaliste. Il est essentiel d'adopter une approche à deux volets concernant, d'une part, l'interdiction et, d'autre part, la restriction ou la réglementation, qui consisterait à interdire les systèmes d'armes fonctionnant sans contrôle humain et à réglementer les autres systèmes, afin de mettre en place l'architecture juridique universelle nécessaire pour créer un environnement permettant de tirer le meilleur parti des nouvelles possibilités offertes par les applications militaires de l'IA tout en relevant les difficultés qui se posent de manière réaliste, efficace et opportune.

Le paysage politique international entourant l'IA dans le domaine militaire est loin d'être unifié. L'Égypte suit de près les multiples initiatives internationales en la matière, qui témoignent d'une prise de conscience croissante des risques connexes. Néanmoins, les délibérations menées dans le cadre de ces initiatives ont révélé des divergences de vues, de perception des menaces et de priorités, et il nous faut donc lancer une mise en garde contre le risque de voir se créer un cadre politique fragmenté

ou des mécanismes concurrents, comme cela a été le cas dans d'autres domaines liés aux technologies nouvelles et émergentes.

Il est manifestement nécessaire de rationaliser ces initiatives et de les placer sous les auspices de l'Organisation des Nations Unies par souci d'inclusivité et d'efficacité. L'Organisation des Nations Unies et ses mécanismes de désarmement forment la seule enceinte efficace et inclusive qui puisse permettre d'élaborer les règles et les normes internationales nécessaires dans ce domaine, d'autant que le rythme auquel évolue la technologie continue de devancer très largement celui de la réglementation internationale.

Il est donc impératif de mettre en place une plateforme universelle, indépendante, unique et fiable sous les auspices de l'Organisation pour discuter de la future gouvernance de l'IA dans le domaine militaire. Le mécanisme envisagé sous l'égide de l'ONU doit être conçu de manière à éviter certaines dichotomies contreproductives qui se font jour, notamment, la dichotomie entre les efforts légitimes visant à garantir le respect de la loi et de l'éthique et la tendance à promouvoir les intérêts militaires sans tenir compte des incidences humanitaires.

Il convient également de souligner que, bien qu'il faille saluer les discussions menées au sein du Groupe d'experts gouvernementaux sur les armes létales autonomes dans le cadre de la Convention sur l'interdiction ou la limitation de l'emploi de certaines armes classiques qui peuvent être considérées comme produisant des effets traumatiques excessifs ou comme frappant sans discrimination, cette plateforme ne peut pas se substituer au processus qu'il est prévu de mener au sein de l'Organisation des Nations Unies sur les applications de l'IA dans le domaine militaire, étant donné que le Groupe d'experts n'est pas universel par nature et qu'il n'a pas non plus pour mandat de traiter un sujet d'une telle versatilité et d'une telle diversité. Il est également regrettable que les progrès réalisés au sein du Groupe d'experts restent minimes et qu'aucun résultat tangible n'ait encore été obtenu.

Les technologies de l'IA présentent, certes, des possibilités, mais également une série de risques inhérents à leurs caractéristiques, car elles peuvent fonctionner de manière imprévisible et inexplicable. Au nombre de ces risques figurent la désinformation, l'escalade involontaire et les risques cybernétiques, ainsi que l'utilisation inappropriée et la prolifération au profit d'acteurs non étatiques. Les risques peuvent être nouveaux ou venir compliquer ceux qui existent déjà.

Il est largement admis qu'il existe un large éventail d'applications militaires envisageables de l'IA. Toutefois, des mesures significatives visant à élaborer leur future gouvernance permettront de fixer le juste ordre des priorités en ce qui concerne le risque inhérent à ces applications et leur incidence sur la paix et la sécurité. Cela permet de garantir des discussions ciblées et structurées, tout en évitant les distractions inutiles. Cela dit, l'Égypte est fermement convaincue que l'accent doit être mis non seulement sur la question des systèmes d'armes létaux autonomes, mais également sur d'autres capacités des systèmes autonomes ou semi-autonomes qui permettent le recours à la force et/ou abaissent le seuil de recours à la force, ce qui pourrait entraîner une nouvelle dynamique de course aux armements s'étendant à la fois aux armes conventionnelles et aux armes non conventionnelles. Le potentiel d'autonomie accrue des armes nucléaires et des armes conventionnelles perfectionnées, telles que les missiles hypersoniques, créerait des risques inconnus et transformerait l'avenir des conflits de manière imprévisible.

Il convient également de mettre l'accent sur les activités de commandement et de contrôle et de sélection des cibles, plutôt que sur la planification logistique et le renseignement, la surveillance et la reconnaissance, étant donné que ces activités ont moins d'effets perturbateurs. De même, l'accent doit être mis davantage sur les capacités offensives que sur les capacités défensives.

Les délibérations qu'il est envisagé d'organiser dans le cadre du mécanisme voulu, sous les auspices de l'ONU, viseront tout d'abord à parvenir à une compréhension commune des principaux éléments qui sous-tendent le développement, le déploiement et l'utilisation de l'intelligence artificielle dans le domaine militaire. Ces éléments sont notamment les suivants :

- Le respect absolu du droit international applicable, notamment des principes cardinaux du droit international humanitaire tels que la nécessité, la proportionnalité et la distinction, ainsi que d'autres considérations éthiques, tout au long du cycle de vie et des étapes des applications de l'IA dans le domaine militaire.

- La nécessité de préserver l'élément humain tout au long du cycle de vie des applications militaires de l'IA, notamment le jugement, l'intervention, le contrôle et la surveillance par l'être humain, qui sont les principaux outils permettant de maintenir l'application du principe de responsabilité. Il est nécessaire de veiller à ce que tous les logiciels, algorithmes et projets reposant sur l'utilisation d'applications d'IA dans le domaine militaire continuent d'être passés en revue par des êtres humains et assujettis au principe d'explicabilité. Alors que les États affirment que le contrôle humain sur les systèmes dotés d'IA est maintenu pour des raisons doctrinales, certains pourraient être tentés de rendre leurs systèmes d'armes de plus en plus autonomes afin de servir leurs intérêts militaires.

- L'équilibre entre la réduction des risques de prolifération pour les acteurs non étatiques et la lutte contre les utilisations malveillantes, d'une part, et le maintien du droit des États à acquérir des technologies d'IA et à double usage, d'autre part. Il est essentiel d'éviter d'introduire des mécanismes de supervision internationaux arbitraires ou d'imposer d'éventuels contrôles à l'exportation discriminatoires.

- Le renforcement des capacités, dans le but de garantir un investissement adéquat dans le capital humain, le transfert de technologies et le partage des connaissances et des meilleures pratiques, de manière à préserver le droit des pays en développement à bénéficier des avantages potentiels des diverses applications de l'IA à des fins militaires, et dans le but de réduire la fracture numérique.

- Les limites de l'IA dans le domaine militaire et son interaction avec d'autres technologies nouvelles et émergentes. Il est opportun d'examiner les moyens d'assurer la complémentarité avec d'autres processus menés par l'Organisation des Nations Unies, notamment le Groupe de travail à composition non limitée sur la sécurité du numérique et de son utilisation, compte tenu, par exemple, des liens entre l'IA et les cyberopérations. En outre, les débats porteront principalement sur le domaine militaire, à l'exclusion des domaines de sécurité plus larges.

Enfin, il importe de veiller à l'inclusivité et à l'équité dans l'élaboration des voies de gouvernance pour une IA responsable, tenue de rendre des comptes et centrée sur l'être humain au sein des Nations Unies. Les perspectives des parties prenantes multiples fournissent des contributions clés qui alimentent les débats. Toutefois, leur participation est sans préjudice de la prérogative souveraine des États dans le processus d'élaboration des politiques.

El Salvador

[Original : espagnol]
[10 avril 2025]

Contexte

Ces dernières années, l'utilisation de l'intelligence artificielle (IA) dans le domaine militaire a joué un rôle majeur. De nombreux rapports indiquent que ces nouvelles technologies deviennent de plus en plus sophistiquées et omniprésentes, ces outils informatiques étant utilisés dans la planification militaire et les processus de prise de décision, y compris pour déterminer qui ou quoi attaquer. L'utilisation de ces technologies soulève de nombreuses questions sur les conséquences globales, les incidences juridiques et les risques pour la population civile. Le débat sur les retombées de ces technologies pour les négociations multilatérales portant sur les systèmes d'armes autonomes d'un point de vue politique, juridique et humanitaire en est un exemple. Il est toutefois admis que les applications militaires de l'IA sont beaucoup plus vastes.

Nous devons donc mieux comprendre l'utilisation et les applications de l'IA dans le contexte militaire, en particulier pour les tâches liées au ciblage militaire et à l'emploi de la force.

La question de l'utilisation responsable de l'IA dans le domaine militaire est devenue particulièrement remarquable à la suite des discussions qui ont eu lieu durant le premier Sommet sur l'intelligence artificielle responsable dans le domaine militaire, qui s'est tenu au Royaume des Pays-Bas en février 2023. De plus, cette question a commencé à prendre de l'importance lors des sessions du Groupe d'experts gouvernementaux sur les technologies émergentes dans le domaine des systèmes d'armes létaux autonomes, basé à Genève.

Il est intéressant de mentionner que, jusqu'à présent, les applications et les utilisations de l'IA ont été principalement abordées pendant les débats portant sur les systèmes d'armes autonomes. Toutefois, l'utilisation de l'IA à des fins militaires est beaucoup plus large et prend une nouvelle dimension, en particulier pour ce qui concerne l'automatisation de certaines fonctions militaires, l'utilisation de l'IA n'étant pas restreinte à l'autonomisation des systèmes d'armes.

D'une manière générale, le débat sur l'IA est un sujet nouveau qui fait encore l'objet de recherches et progresse très rapidement, à tel point que des initiatives sont lancées aux niveaux national, régional et multilatéral pour faire face à ses retombées. Il apparaît clairement que les pays d'Amérique latine et des Caraïbes ne disposent pas du même niveau en matière de technologies et de renforcement des capacités que les pays développés pour repérer et comprendre les possibilités et les difficultés liées à l'utilisation de l'IA. Il est donc important de définir une position nationale qui leur permette de participer activement aux discussions qui s'ouvrent dans les espaces et forums internationaux, et de parvenir ainsi à une coopération relative au renforcement des capacités et aux éléments qui permettront à ces pays de jouer un rôle moteur sur cette question et de comprendre les opportunités et les risques éventuels en termes de sécurité aux niveaux national, régional et mondial.

Initiatives auxquelles El Salvador a participé

- El Salvador a participé au Sommet sur l'intelligence artificielle responsable dans le domaine militaire, qui s'est tenu en 2023 au Royaume des Pays-Bas, et s'est associé à la déclaration qui en a résulté (février 2023).

- Il a également pris part à la conférence latino-américaine et caribéenne sur l'impact social et humanitaire des systèmes d'armes autonomes, au cours de laquelle a été adopté le communiqué de Belém (février 2023).
- Le Salvador fait partie du « Groupe des 16 » dans le cadre des débats du Groupe d'experts gouvernementaux sur les systèmes d'armes létaux autonomes. Il s'agit d'une question à part entière, mais elle est liée à l'utilisation de l'IA dans le domaine militaire.

Position du pays

- Certaines utilisations de l'IA peuvent présenter des avantages dans le domaine militaire, en particulier les utilisations liées à d'autres tâches administratives, telles que l'analyse de données et l'apprentissage automatique sans rapport avec l'intervention humaine dans les opérations militaires. Il s'agit donc d'utilisations qui ne sont pas associées aux fonctions de repérage et de reconnaissance des cibles militaires et plus particulièrement à l'emploi de la force, celles-ci présentant un risque pour la population civile.
- Toutefois, l'utilisation abusive de ces applications peut avoir des effets négatifs, notamment en ce qui concerne la protection des civils et des infrastructures civiles, qui font l'objet de catégories de protection spéciales fondées sur les règles du droit international, notamment le droit international humanitaire et le droit international des droits humains.
- Il importe d'adopter une approche fondée sur les risques qui permette de réglementer et d'interdire certaines fonctions de l'IA, en particulier celles qui limitent un contrôle humain marqué de l'emploi de la force, celles qui perpétuent les biais algorithmiques en raison de l'utilisation de bases de données non représentatives ou de données anciennes et qui présentent des risques en matière de droits humains, et celles qui, à long terme, mettent en péril la sécurité internationale, en particulier lorsqu'une machine détient le pouvoir de décider de la vie et de la mort d'un être humain, ou lorsque ces outils sont dotés d'éléments technologiques très sophistiqués, tels que l'auto-apprentissage, ce qui pourrait avoir de graves incidences humanitaires, sociales, économiques, politiques et même environnementales.
- Il est désormais urgent de mettre en place une réglementation adéquate dans le domaine de l'IA, car il importe de garantir son développement sûr et éthique, ce qui permettra de protéger les utilisateurs et la société contre les abus et les risques, et de favoriser l'innovation en offrant un environnement clair et sûr aux développeurs et aux chercheurs.
- L'objectif ultime est l'élaboration d'instruments juridiquement contraignants, mais nous considérons que les progrès de ces technologies sont plus rapides que l'évolution ou l'élaboration du droit international dans ce domaine, et pensons donc qu'il faut conserver une approche axée sur un comportement responsable, qui pourra servir de socles à des engagements juridiques d'ensemble permettant de mieux examiner cette question.
- Il est important d'examiner le défi que les technologies émergentes imposent en matière de sécurité. C'est notamment le cas de l'utilisation des technologies des matériaux, telles que l'impression 3D, pour la fabrication d'armes légères et de petit calibre, de l'utilisation de la robotique pour le développement de robots dotés de capacités autonomes dans le domaine militaire, et de certaines utilisations et applications de l'IA qui, en raison de leur double usage, peuvent reproduire des biais dans les fonctions de commandement et de contrôle en cas de conflit armé, ce qui induit un risque accru pour les civils.

- La perte ou le remplacement du contrôle militaire peut entraîner des risques involontaires. L'IA peut accroître les capacités humaines, mais l'absence de contrôle dans le contexte militaire peut présenter d'autres risques, qui doivent être étudiés de manière approfondie. L'assistance apportée par l'IA dans le domaine militaire devrait renforcer ou éclairer la prise de décision dans des contextes particuliers, mais jamais remplacer la prise de décision et le raisonnement par des humains.
- L'utilisation militaire de l'IA doit respecter le droit international, le droit international des droits humains et le droit international humanitaire, et servir l'intérêt général.
- Les capacités des pays doivent être renforcées pour qu'ils puissent mettre au jour les risques liés à l'utilisation abusive de l'IA et son lien avec le droit international.
- D'autres acteurs qui font partie du processus de création et de développement de technologies de ce type, comme les entreprises et les universités, entre autres, devraient participer aux discussions multilatérales, et la coopération internationale entre les parties concernées devrait être encouragée afin de tirer parti des avantages qu'offre l'IA lorsqu'elle est utilisée à des fins pacifiques en faveur du développement des pays.

Espagne

[Original : espagnol]
[11 avril 2025]

Introduction

L'intelligence artificielle (IA) est une révolution dans tous les domaines, y compris celui de la sécurité et de la défense. Son développement et son intégration sont porteurs de progrès et d'opportunités considérables, tout en créant de nombreux défis.

L'adoption de cette technologie dans les forces armées ne redéfinit pas seulement la manière dont les opérations militaires sont menées, mais transforme également l'équilibre stratégique global. Le développement et l'utilisation de l'IA au Ministère de la défense reposent sur un déploiement militaire responsable, éthique et légitime, dans le respect du droit international humanitaire et des droits humains.

L'IA modifie la conception traditionnelle de la puissance et de la sécurité militaires, en fournissant des capacités avancées pour la collecte et l'analyse de données, la prise de décision et l'exécution d'opérations dans des environnements multisectoriels. Il s'agit d'un changement de paradigme dans la manière dont les États abordent la défense et la sécurité, ce qui permet de répondre de façon plus rapide et précise aux menaces émergentes.

Dans le domaine militaire, l'IA a des effets perturbateurs sur le champ de bataille, lui-même imprévisible, entraînant un changement de paradigme dans la planification et la conduite des opérations militaires. Elle influe aussi d'autres questions relevant du domaine militaire : la logistique, la formation, la gestion et l'interprétation de l'information, le renseignement, la surveillance, l'acquisition d'objectifs et la reconnaissance.

En cohérence avec l'engagement de l'Espagne en faveur d'une utilisation responsable de l'IA, il convient de souligner l'organisation du Sommet sur l'intelligence artificielle responsable dans le domaine militaire de 2025 en tant que

pays hôte, en plus de son adhésion à l'appel à l'action (La Haye, 2023) et au plan d'action présenté lors du dernier Sommet en 2024.

Cadre conceptuel et réglementaire adopté par le Ministère de la défense

Le développement, le déploiement et l'application de l'IA au Ministère de la défense sont guidés par un ensemble de principes fondamentaux qui garantissent une utilisation sûre, éthique et conforme aux réglementations nationales et internationales. Ces principes, qui sont compilés dans la stratégie de développement, de mise en œuvre et d'utilisation de l'intelligence artificielle au Ministère de la défense (« Estrategia de desarrollo, implantación y uso de la Inteligencia Artificial en el Ministerio de Defensa », élaborée comme suite à la résolution 11197/2023 du Secrétaire d'État à la défense) et alignés sur la stratégie adoptée par l'Organisation du Traité de l'Atlantique Nord (OTAN) en matière d'IA (en 2021 et révisée en 2024), visent à maximiser les possibilités offertes par l'IA pour la défense, tout en cherchant à atténuer les risques associés à son utilisation dans le domaine militaire, comme indiqué ci-dessous.

- Légalité : les applications basées sur l'IA seront développées et utilisées dans le respect du droit national et international applicable, notamment la Déclaration universelle des droits de l'homme et le droit international humanitaire.
- Responsabilité humaine et application du principe de responsabilité : tout développement et utilisation de l'IA doivent faire clairement intervenir une supervision humaine afin de garantir l'obligation de rendre des comptes et l'établissement des responsabilités.
- Intelligibilité et traçabilité : les applications basées sur l'IA seront compréhensibles et transparentes pour le personnel concerné, notamment grâce à l'utilisation de méthodologies, de sources et de procédures vérifiables.
- Fiabilité et transparence : les applications basées sur l'IA seront axées sur des cas d'utilisation explicites, bien définis et délimités, et des informations seront fournies pour favoriser une compréhension générale de ces applications par toutes les parties prenantes. La sûreté, la sécurité et la robustesse de ces capacités feront l'objet d'essais et de garanties dans le cadre de ces cas d'utilisation tout au long du cycle de vie.
- Gouvernance : les applications basées sur l'IA seront élaborées et déployées dans le respect des fonctions prévues, ce qui permettra de détecter et d'éviter les conséquences imprévues. Des mécanismes de déconnexion ou de désactivation devront être activés lorsqu'un comportement imprévu ou indésirable est détecté.
- Atténuation des biais : toutes les mesures nécessaires seront prises pour minimiser les erreurs et les orientations subjectives dans le développement et les utilisations de l'IA.
- Vie privée : le développement, le déploiement et l'utilisation d'applications basées sur l'IA doivent respecter la vie privée des personnes, dès la conception et tout au long du cycle de vie.

En ce qui concerne le cadre réglementaire, pour le développement, la mise en œuvre et l'utilisation de l'IA dans le domaine militaire au Ministère de la défense, un ensemble de normes et de bonnes pratiques est en cours d'élaboration afin de garantir une utilisation responsable, efficace et conforme aux cadres juridiques nationaux et internationaux, et de veiller en particulier au strict respect du droit international humanitaire et des droits humains.

Possibilités

Le Ministère de la défense concentre le développement de ses capacités d'IA dans de multiples domaines afin d'améliorer l'efficacité des forces armées. Selon la stratégie, l'IA est principalement utilisée pour les opérations, le renseignement, la logistique et la cybersécurité, ainsi que pour l'aide à la décision.

L'IA permettra d'améliorer la précision, la rapidité et l'efficacité de la prise de décision au cours d'une opération militaire, dans le respect constant du droit international humanitaire, ce qui permettra d'exécuter les missions plus efficacement et de réduire les risques pour les troupes, tout en contribuant à renforcer la protection des civils et des biens de caractère civil dans les conflits armés.

Sa capacité d'analyser de grands volumes de données en temps réel améliore la perception de la situation et les capacités de réaction aux menaces, ce qui renforce la sécurité opérationnelle. Dans toutes ces améliorations, le contrôle humain est toujours présent et aucune responsabilité n'est déléguée aux machines.

Pour ce qui concerne la formation et l'entraînement militaires, dans le cadre du groupe des commandants du C5 (Royaume-Uni, France, Allemagne, Italie et Espagne), des travaux sont en cours pour créer un espace de collaboration sur l'IA dans l'enseignement militaire.

L'Espagne collabore aussi avec le Comité de surveillance Données et intelligence artificielle de l'OTAN sur l'utilisation responsable des données et de l'IA dans le domaine militaire.

En outre, le Ministère de la défense a annoncé des investissements stratégiques dans certaines régions afin d'encourager les projets liés à l'IA et à d'autres technologies de pointe. Ces investissements visent non seulement à renforcer l'industrie, mais aussi à promouvoir la revitalisation industrielle de nouvelles zones régionales.

Difficultés

Le développement et l'application de l'IA dans le domaine militaire doivent être alignés sur les cadres normatifs nationaux et internationaux, y compris le respect du droit international humanitaire. Des travaux sont donc en cours pour garantir un contrôle humain efficace sur les décisions critiques associées à l'utilisation de l'IA dans les opérations militaires.

D'autre part, pour ce qui concerne la protection de la vie privée et des données, la collecte et le traitement massifs de données pour former des modèles d'IA présentent des risques en termes de protection des données personnelles et de sécurité de l'information.

Sûreté et fiabilité

Le principal défi concerne l'utilisation sûre et fiable de l'IA, les principaux risques associés étant les suivants :

- Les données d'apprentissage utilisées pour les algorithmes peuvent présenter des biais, ce qui peut donner lieu à des décisions erronées ou engendrer des conséquences involontaires.
- Un mauvais entraînement des modèles d'IA peut conduire à des interprétations erronées, qui peuvent à leur tour avoir des répercussions potentiellement catastrophiques dans les opérations militaires.
- Les systèmes d'IA peuvent être la cible de cyberattaques, susceptibles de manipuler leur comportement ou les rendre inutilisables.

- Il existe un risque d'empoisonnement des données, si des acteurs malveillants modifient les ensembles de données d'entraînement pour créer des erreurs dans les algorithmes.

En Espagne, le développement de l'IA dans le domaine militaire est régi par les principes de responsabilité et de surveillance continue, des mécanismes d'évaluation des risques, d'audit et de traçabilité étant prévus à chaque phase du cycle de vie du système. Tout développement de l'IA, ainsi que son utilisation, devra permettre une supervision humaine claire afin qu'il soit possible de rendre compte et d'assigner des responsabilités comme il se doit : les activités humaines liées aux performances de l'IA devront être traçables, en parallèle du fonctionnement de l'IA, et la décision finale ne devra pas être laissée aux machines.

L'IA doit aussi être fiable et prévisible, tout en conservant un niveau d'autonomie contrôlé et surveillé par des opérateurs formés.

Toute solution utilisant l'IA sera examinée dans un environnement différent de celui dans lequel elle a été entraînée et sera soumise à des tests non fonctionnels : des tests de charge, de stress et de performance seront menés dans le cadre de scénarios changeants définis, afin d'étudier son comportement et l'écart admissible.

En outre, ces capacités d'IA seront soumises à des tests rigoureux et à des audits constants tout au long de leur cycle de vie, ce qui permettra de détecter rapidement les erreurs potentielles et d'améliorer leur fiabilité opérationnelle. Des protocoles de surveillance et de contrôle humains seront mis en œuvre à tous les stades du déploiement, afin de faire en sorte que les décisions critiques ne sont pas déléguées exclusivement à l'IA. À cet égard, des travaux sont en cours pour que les évolutions de l'IA soient certifiées par des organismes reconnus.

Afin d'améliorer la robustesse des systèmes basés sur l'IA et de les protéger des agissements extérieurs, il faut intégrer la sécurité dans la conception, en veillant à ce que ces systèmes soient résistants aux cyberattaques et aux manipulations adverses, pour garantir l'intégrité des données et des modèles utilisés.

L'IA peut être la cible d'attaques telles que l'empoisonnement de données ou la manipulation de modèles, ce qui nécessite une surveillance continue des performances du système et la conduite de tests de validation et d'audits réguliers. L'élaboration de plans de sauvegarde et de basculement sera encouragée, afin de garantir l'opérabilité des systèmes en cas de scénarios défavorables.

La collaboration avec les agences de cybersécurité et les experts en IA est également encouragée, afin que les forces armées disposent des meilleurs outils et stratégies pour protéger ces systèmes contre les menaces extérieures et favoriser leur fiabilité opérationnelle.

Par ailleurs, le talent du personnel et la formation de celui-ci à ces technologies sont essentiels et constituent l'un des quatre piliers sur lesquels le Ministère de la défense se concentre, en veillant à ce que les opérateurs comprennent la portée et les limites de ces systèmes et puissent intervenir en cas de déviation de leur comportement. La formation et la sensibilisation du personnel à l'utilisation légale et éthique de l'IA sont indispensables pour atténuer les risques liés aux préjugés et garantir que son utilisation dans les forces armées est objective, fiable et conforme aux réglementations nationales et internationales, en particulier au droit international humanitaire.

Des travaux sont actuellement menés pour élaborer un guide des meilleures pratiques qui pourrait constituer la base d'un document auquel contribueraient tous les services du Ministère de la défense. Les bonnes pratiques proposées par l'OTAN pour une utilisation responsable de l'IA dans le domaine militaire ont été diffusées.

Un exemple est l'évaluation et la boîte à outils « IA responsable » de l'OTAN (RAI, pour « Responsible Artificial Intelligence »), qui vise à rendre opérationnels les principes de l'utilisation responsable de l'IA adoptés par l'OTAN, dont la légalité, la responsabilité, la traçabilité, la fiabilité, la gouvernance et l'atténuation des biais.

Fédération de Russie

[Original : russe]
[10 avril 2025]

La Fédération de Russie se félicite de l'adoption de la résolution 79/239 de l'Assemblée générale en date du 24 décembre 2024 et, en application du paragraphe 7 de ladite résolution, a l'honneur de soumettre sa contribution nationale au rapport que le Secrétaire général présentera à la quatre-vingtième session de l'Assemblée, dans la perspective de futurs débats entre les États Membres.

Introduction

La Fédération de Russie attache une grande importance à l'application de l'intelligence artificielle dans le domaine militaire. Nous souhaitons que cette question continue de faire l'objet d'un débat approfondi dans les instances internationales compétentes.

Nous considérons que le Groupe d'experts gouvernementaux sur les systèmes d'armes létaux autonomes, créé en vertu de la Convention sur certaines armes classiques, est la plateforme optimale pour ce débat. Il appartient au Groupe de maintenir un équilibre raisonnable entre les préoccupations humanitaires et les intérêts légitimes liés à la défense exprimés par les États en ce qui concerne ces équipements, et de prendre des décisions par consensus. L'examen, par le Groupe, des applications militaires de l'intelligence artificielle a une large portée et ne se limite pas à la question des systèmes d'armes létaux autonomes mais touche à plusieurs aspects importants (juridiques, techniques, militaires) liés à l'utilisation de cette technologie à des fins militaires.

Nous prenons note de l'examen de cette question dans le cadre des structures relatives à la maîtrise des armements, au désarmement et à la non-prolifération. Ces travaux visent à analyser les risques liés à l'intelligence artificielle et les possibilités qu'elle offre en ce qui concerne le respect par les États Parties des obligations découlant des instruments juridiques internationaux pertinents.

Nous nous félicitons que les États Membres soient prêts à discuter des applications militaires de l'intelligence artificielle au sein de la Commission du désarmement, dans le cadre du débat sur les technologies émergentes dans le contexte de la sécurité internationale. Cet échange de vues vise à convenir de recommandations sur les aspects de l'intelligence artificielle « militaire » qui ne sont pas examinés dans d'autres instances.

Au cours de leurs travaux, les forums internationaux susmentionnés doivent mettre l'accent sur l'élaboration d'une terminologie commune, l'application du droit international, le contrôle humain, le respect du principe de responsabilité ainsi que les risques liés à cette technologie et les possibilités qu'elle offre.

Définition

L'absence, en droit international, d'une définition consensuelle des systèmes d'armes et du matériel militaire basés sur l'intelligence artificielle complique l'examen de cette question. L'élaboration d'une définition commune de ces équipements et, plus généralement, de la terminologie associée aux applications

militaires de cette technologie donnera une plus grande clarté à ce sujet et aux futurs débats y relatifs.

Cette définition pratique doit satisfaire aux prescriptions suivantes :

a) Elle doit décrire les types de systèmes d'armes et de matériel militaire basés sur l'intelligence artificielle et les principales caractéristiques de leur utilisation ;

b) Elle ne doit pas se limiter à la définition actuelle de ces équipements mais tenir compte de la manière dont ils pourraient évoluer ;

c) Elle doit être universellement comprise par la communauté d'experts (chercheurs, ingénieurs, techniciens, militaires, juristes et éthiciens) ;

d) Elle ne doit pas être vue comme une limite au progrès technologique ou un frein à la recherche dans les domaines de la robotique civile et de l'intelligence artificielle ;

e) Elle ne doit pas définir les systèmes d'armes et le matériel militaire basés sur l'intelligence artificielle uniquement en décrivant leurs fonctions.

Il convient d'éviter de classifier ces équipements en faisant la distinction entre les « mauvais » équipements et les « bons » équipements, c'est-à-dire de les classer en fonction des préférences politiques d'un groupe d'États en particulier.

Les systèmes militaires hautement automatisés ne doivent pas être rangés dans une catégorie « spéciale » pour laquelle des restrictions et des interdictions immédiates sont nécessaires. C'est en effet grâce à ce niveau d'automatisation que de tels systèmes fonctionnent bien dans des situations de combat dynamiques et dans différents environnements et qu'ils sont suffisamment sélectifs et précis, ce qui garantit leur conformité avec les principes et les normes du droit international, notamment le droit international humanitaire.

Systèmes d'armes et matériel militaire basés sur l'intelligence artificielle dans le contexte du droit international

Il est universellement admis que le droit international, notamment le droit international humanitaire, s'applique pleinement aux systèmes d'armes basés sur l'intelligence artificielle.

La Fédération de Russie estime qu'il n'existe actuellement aucune raison valable d'imposer de nouvelles restrictions ou interdictions relatives aux systèmes d'armes basés sur l'intelligence artificielle ni d'actualiser ou d'adapter le droit international, notamment le droit international humanitaire, s'agissant de ces systèmes. Les débats portant sur l'adoption de « règles de conduite » ou de normes et principes d'utilisation « responsable » des systèmes d'armes et du matériel militaire basés sur l'intelligence artificielle sont également prématurés. Le concept d'application « responsable » de l'intelligence artificielle prôné par les pays occidentaux repose sur des critères très controversés ne relevant pas du droit international (notamment du droit international humanitaire), soulève de nombreuses questions et ne recueille pas le consensus de la communauté internationale.

Les principes d'humanité, la conscience publique et les droits humains ne sauraient à eux seuls justifier de façon inconditionnelle l'imposition de restrictions et d'interdictions pour certains types d'armes et de matériel militaire. Les inquiétudes concernant les systèmes d'armes et le matériel militaire basés sur l'intelligence artificielle doivent être apaisées grâce à l'application, de bonne foi, des normes juridiques internationales déjà en vigueur.

Le strict respect des normes et des principes du droit international, notamment du droit international humanitaire, dans les situations de conflit armé demeure l'une des priorités de la Fédération de Russie. Les forces armées de la Fédération de Russie se conforment strictement aux normes du droit international humanitaire inscrites dans les actes juridiques fédéraux et ministériels. Dans les règlements et les programmes de formation visant toutes les catégories de militaires figurent des questions relatives au respect du droit international humanitaire, notamment en ce qui concerne l'utilisation de nouveaux types d'armes. En 2022 a été adopté un document de réflexion des forces armées de la Fédération de Russie sur le développement et l'utilisation de systèmes d'armes basés sur l'intelligence artificielle.

La législation russe est pleinement conforme aux principes directeurs sur les systèmes d'armes basés sur l'intelligence artificielle, approuvés par consensus en 2019 par les États Parties à la Convention sur certaines armes classiques. Nous considérons que la poursuite de l'échange d'informations sur les mesures concrètes de mise en œuvre de ces principes directeurs au niveau national permettra d'améliorer la confiance et la transparence.

Contrôle des systèmes d'armes et du matériel militaire basés sur l'intelligence artificielle

Nous considérons que le contrôle humain du fonctionnement des systèmes d'armes et du matériel militaire basés sur l'intelligence artificielle est une limitation importante. Un opérateur ou un système de gestion de niveau supérieur doit faire partie de la boucle de contrôle de ces systèmes pour pouvoir en modifier le fonctionnement, et notamment les désactiver complètement ou partiellement.

La Fédération de Russie part du principe qu'un être humain porte toujours la responsabilité de la décision d'employer la force. Le contrôle est fondé sur toutes les informations disponibles au moment de la prise de décision. Toutefois, les États doivent pouvoir décider des formes et des méthodes concrètes de ce contrôle humain, qui ne doit pas forcément être direct.

On peut assurer le contrôle des systèmes d'armes et du matériel militaire basés sur l'intelligence artificielle de la manière suivante :

a) En améliorant leur fiabilité et leur tolérance aux pannes ;

b) En limitant les types de cibles ;

c) En limitant leur durée de fonctionnement, leur couverture géographique et leur échelle d'application ;

d) En effectuant des interventions et en désactivant ces systèmes en temps utile ;

e) En les mettant à l'essai dans des environnements opérationnels réalistes ;

f) En autorisant les personnes maîtrisant les procédures d'utilisation de ces équipements à les manier (contrôler) ;

g) En surveillant la production d'éléments individuels et de la pièce dans son ensemble ;

h) En surveillant le démantèlement et la destruction des éléments individuels et de la pièce dans son ensemble.

Nous considérons qu'il est inutile d'évoquer, dans ce débat, les notions de « contrôle humain véritable », « forme et degré d'intervention humaine », « évaluation et contrôle humains répondant au contexte » et « prévisibilité, fiabilité,

traçabilité et explicabilité » avancées par certains États, étant donné qu'elles n'ont généralement aucun rapport avec le droit et ne font que politiser le débat.

Principe de responsabilité

La Fédération de Russie estime que les États et les personnes (notamment les développeurs et les fabricants) portent à tout moment la responsabilité, conformément au droit international, de leur décision de créer et d'utiliser des systèmes d'armes et du matériel militaire basés sur l'intelligence artificielle. La responsabilité de l'utilisation de ces équipements incombe à la personne chargée de leur assigner une tâche et d'ordonner leur utilisation. La personne qui utilise des systèmes d'armes et du matériel militaire basés sur l'intelligence artificielle doit posséder les connaissances et les compétences nécessaires pour les faire fonctionner et les utiliser, et décider du bien-fondé, des formes et des méthodes d'utilisation de ces équipements.

Possibilités et limites associées aux systèmes d'armes et au matériel militaire basés sur l'intelligence artificielle

Il est de notoriété publique que les systèmes d'armes et le matériel militaire basés sur l'intelligence artificielle peuvent mieux effectuer des tâches que les opérateurs humains et réduire le risque d'erreurs. Ces équipements peuvent notamment réduire considérablement les effets néfastes, sur le droit international (notamment le droit international humanitaire), liés aux erreurs, à l'état mental et physique ainsi qu'aux convictions morales, religieuses ou éthiques de l'opérateur. Leur utilisation peut améliorer la précision des armes dirigées contre des objectifs militaires et contribuer à diminuer le risque de frappes involontaires contre des civils et des biens de caractère civil.

L'évaluation des risques potentiels liés à l'utilisation de systèmes d'armes et de matériel militaire basés sur l'intelligence artificielle ainsi que les mesures d'atténuation de ces risques doivent faire partie du cycle de conception, de développement, de mise à l'essai et de déploiement de nouvelles technologies dans tout système militaire.

On pourrait réduire les risques liés à ces équipements de la manière suivante :

a) En gérant efficacement le cycle de vie ;

b) En procédant à des essais complets à tous les stades du cycle de vie, notamment dans des conditions proches de la réalité ;

c) En assurant la fiabilité de ces équipements et leur tolérance aux pannes ;

d) En définissant des critères de disponibilité opérationnelle ;

e) En garantissant une protection maximale contre les accès non autorisés ;

f) En formant les opérateurs ;

g) En recourant en priorité à l'intelligence artificielle pour la collecte et le traitement des informations nécessaires à la prise de décision militaire ;

h) En veillant à ce que l'opérateur assure le contrôle continu des actions de ces systèmes et puisse mettre fin d'urgence à une mission de combat ;

i) En empêchant que ces systèmes tombent entre les mains d'acteurs non étatiques, qui pourraient les utiliser à des fins illicites.

Ces mesures peuvent être prises à tous les stades du cycle de vie (développement, production, exploitation, destruction) des armes et du matériel militaire et spécial.

Voie à suivre

Nous estimons qu'il serait utile que les États poursuivent l'examen des questions liées aux applications militaires de l'intelligence artificielle au sein du Groupe d'experts gouvernementaux sur les systèmes d'armes létaux autonomes, qui est la plateforme internationale optimale pour ce genre de discussion, dans le cadre des structures relatives à la maîtrise des armements, au désarmement et à la non-prolifération, et au sein de la Commission du désarmement. Cela étant, les discussions menées au sein d'une instance ne doivent pas faire double emploi avec l'échange de vues qui a déjà lieu dans des plateformes de dialogue parallèles.

Nous sommes défavorables à la fragmentation de l'action menée dans ce domaine. Il est contre-productif de transférer la question des applications militaires de l'intelligence artificielle à une quelconque autre instance internationale, de créer des plateformes supplémentaires pour examiner cette question ou d'en débattre dans un format restreint sans la participation de la grande majorité des États Membres de l'ONU (notamment les principaux développeurs de systèmes d'armes basés sur l'intelligence artificielle, dont fait partie la Fédération de Russie).

Les discussions menées dans le cadre de sommets non inclusifs sur l'application responsable de l'intelligence artificielle dans le domaine militaire organisés par un groupe d'États occidentaux ainsi que de sommets sur l'intelligence artificielle en général ont des effets destructeurs. Ces activités et les documents adoptés ne tiennent pas compte de l'avis de toutes les parties prenantes et ne sauraient être considérés comme une base pour des travaux ultérieurs devant être fondés sur une compréhension commune de la question. Ils créent des clivages et ne permettent pas de conjuguer les efforts dans ce domaine.

Les tentatives de cimenter les approches unilatérales de la question dans d'autres instances, notamment dans le cadre de ces prétendus sommets, en contournant les plateformes multilatérales compétentes, auront des conséquences extrêmement néfastes. Elles risquent de compromettre gravement les travaux constructifs et inclusifs qui sont menés concernant les applications militaires de l'intelligence artificielle et de diviser l'action visant à élaborer des définitions et des recommandations communes dans ce domaine.

Dans le cadre des discussions menées dans les instances internationales susmentionnées, nous pensons qu'il faut s'employer en particulier à adopter une terminologie et des approches communes en ce qui concerne l'application du droit international, notamment du droit international humanitaire, aux systèmes d'armes et au matériel militaire basés sur l'intelligence artificielle, le contrôle humain de ces équipements ainsi que les risques liés à cette technologie et les possibilités qu'elle offre.

La Fédération de Russie prie le Secrétaire général de tenir compte des propositions présentées ci-dessus dans son rapport de fond, en application du paragraphe 7 de la résolution 79/239 de l'Assemblée générale, et de faire figurer le présent document en annexe de ce rapport.

Finlande

[Original : anglais]
[11 avril 2025]

La Finlande a le plaisir de présenter ses vues concernant la résolution 79/239 de l'Assemblée générale intitulée « L'intelligence artificielle dans le domaine militaire et ses conséquences pour la paix et la sécurité internationales », adoptée le 24 décembre 2024, dans laquelle le Secrétaire général est prié de solliciter les vues

des États Membres « sur les possibilités et les difficultés que l'application de l'intelligence artificielle dans le domaine militaire présente pour la paix et la sécurité internationales, en particulier dans des domaines autres que les systèmes d'armes létaux autonomes ».

L'adoption de principes ou de réglementations internationales sur l'application de l'IA dans le domaine militaire est fondamentale pour garantir le respect du droit international, accroître la sécurité et réduire les risques potentiels de conflit. Dans le même temps, il est nécessaire de favoriser le développement de capacités de défense nationale conformes au droit international. La Finlande s'est engagée à développer, déployer et utiliser des capacités d'IA dans le domaine militaire de manière responsable, conformément au droit international, en particulier au droit international humanitaire, et d'une manière qui ne porte pas atteinte à la paix, à la sécurité et à la stabilité internationales, tout en poursuivant ses efforts en matière de recherche, de développement, d'expérimentation et d'innovation dans le domaine de la technologie de l'IA.

Il est de plus en plus important de recenser les conséquences des technologies de rupture en matière de politique étrangère, de sécurité et de défense et de se doter de moyens pour y faire face. La Finlande participe activement aux débats mondiaux sur la réglementation des technologies, en défendant les droits fondamentaux et les droits humains, ainsi qu'en s'attaquant aux risques connexes, dans le cadre du développement et de l'application de l'IA et des politiques correspondantes.

Il convient de recenser les risques liés aux technologies de rupture et de reconnaître les possibilités qu'elles offrent en matière de sécurité, de renforcement des capacités de défense, de croissance économique, de productivité, de développement durable, de compétences technologiques et d'investissements sectoriels.

Perspectives

Les technologies de rupture offrent des possibilités considérables de faire progresser divers secteurs, de favoriser la transition vers une économie propre, d'encourager une croissance économique durable et d'améliorer l'efficacité et la productivité. Elles sont également susceptibles d'améliorer la sécurité, l'éducation, le bien-être et la santé au niveau mondial.

L'IA et d'autres technologies émergentes offrent des possibilités de faire progresser les capacités de défense tout en façonnant fondamentalement l'avenir des champs de bataille et des moyens et méthodes de guerre. Les progrès technologiques favorisent une collecte d'informations et un traitement des données plus efficace, une meilleure perception de la situation, une prise de décision plus rapide et une collaboration plus précise et de plus grande portée. Les systèmes télécommandés et autonomes sans pilote sont de plus en plus importants dans la guerre moderne, et ils changeront l'avenir de la guerre, des opérations et des champs de bataille. Il sera de plus en plus important d'anticiper les progrès technologiques, d'intégrer les technologies émergentes dans les systèmes de défense et de tirer parti de l'inattendu à mesure que le rythme du développement technologique s'accélèrera. Le progrès technologique peut également compenser l'infériorité numérique.

Difficultés

Dans le même temps, il importe d'établir une large compréhension des menaces pour la sécurité, des potentielles utilisations abusives, des questions relatives aux droits humains et des interdépendances liées au développement de technologies de rupture telles que l'intelligence artificielle. Au fur et à mesure de leur développement, ces technologies soulèveront de nouvelles difficultés pour les secteurs de la défense

et de la sécurité, en particulier. Le développement de l'IA rend les cyberattaques, les activités d'influence de l'information et la désinformation, qui en est l'un des instruments, plus ciblées et plus efficaces. En outre, l'IA est déjà utilisée pour influencer les élections. Dans un tel environnement, il convient également d'accorder davantage d'attention à la sécurité des informations confidentielles.

Le droit international, en particulier la Charte des Nations Unies, le droit international des droits humains et le droit international humanitaire, s'applique pleinement au cyberespace. Le respect du cadre des Nations Unies régissant le comportement responsable des États dans le cyberespace et l'adhésion à ce cadre restent essentiels au maintien de la paix, de la sécurité et de la stabilité internationales. Le développement technologique soulève de nouvelles questions, par exemple, le cyberenvironnement, l'utilisation de l'IA, les nouvelles technologies de l'armement et l'exploitation des matières premières essentielles. Les activités d'influence hybrides peuvent consister en des pratiques visant à entraver l'obligation effective de rendre des comptes en vertu du droit international. La Finlande préconise de prendre fortement en compte les droits fondamentaux et les droits humains, ainsi que les risques qui y sont liés, lors du développement et de l'application de l'IA et de l'élaboration de la réglementation correspondante. Il importe d'établir des principes, des normes et des standards, des politiques et des cadres nationaux pour garantir des applications responsables de l'IA dans le domaine militaire, dans le respect du droit international.

Les avancées technologiques ont offert aux acteurs hostiles de nouvelles possibilités de s'engager dans des activités d'influence hybrides exercées dans des situations n'atteignant pas le seuil de conflit ouvert. Les cyberopérations hostiles font désormais partie intégrante de la politique de puissance et de la gamme d'instruments disponibles pour les activités d'influence menées par les acteurs étatiques. Les opérations cybernétiques, hybrides et d'information sont également menées dans des conditions ordinaires, ce qui peut également brouiller les frontières entre la guerre et la paix. Malgré la nature de plus en plus technologique de la guerre, les capacités de guerre conventionnelle restent importantes, en particulier dans les conflits à grande échelle et à long terme.

De nombreux pays font l'expérience d'intenses activités d'influence de l'information recourant également à l'IA. L'utilisation préjudiciable de l'information est devenue un élément quotidien de l'influence à large spectre, et la concurrence dans l'environnement de l'information s'est accrue.

L'évolution des infrastructures et des technologies et le nombre croissant d'utilisateurs offrent de plus grandes possibilités d'actions hostiles dans le domaine cybernétique. De nombreux pays doivent constamment faire face à la collecte de renseignements sur les réseaux d'information, au cyberespionnage et aux cyberattaques d'acteurs hostiles qui visent également à avoir une incidence physique sur les infrastructures essentielles. Parallèlement aux acteurs étatiques, les acteurs non étatiques à motivation politique ou dirigés par un État jouent un rôle de plus en plus important en tant qu'orchestrateurs d'activités hostiles.

France

[Original : français]
[11 avril 2025]

I. L'impact de l'intelligence artificielle dans le domaine militaire sur la paix et la sécurité internationales

Des opportunités à exploiter

Aider à la planification et à la prise de décision. Les armées françaises travaillent sur l'exploitation des bases de leurs données en matière d'évènements liés aux munitions et explosifs afin de développer des outils prédictifs sur les menaces potentielles dans une zone définie.

Soutenir l'humain. Le système d'intelligence artificielle (IA) pour la formation des personnels navigants « IA FPN » est mis en avant afin d'améliorer la formation des pilotes français, en analysant les données recueillies des vols ou simulations. L'IA peut également aider l'humain face à une grande quantité de données, comme le système « oreille d'or » qui traite massivement des données acoustiques pour orienter l'attention des opérateurs français sur les seuls signaux à valeur ajoutée.

Contrer nos vulnérabilités dans le domaine des technologies de l'information et de la communication. Les technologies d'IA peuvent servir à la cybersécurité et faire face à la prolifération de fausses informations. Les armées françaises s'appuient sur des systèmes de détection des hypertrucages (*deepfakes*).

Favoriser la mise en œuvre du droit international humanitaire et la protection des personnes et biens protégés. L'IA peut contribuer à la mise en œuvre des principes cardinaux du droit international humanitaire, comme la distinction, la proportionnalité et le principe de précaution. L'IA pourra également agir pour la protection des personnes en aidant au déminage des territoires par moyen de drones munis de capteurs couplés à une IA.

Renforcer la maîtrise des armements. L'IA peut être utilisée pour mieux surveiller et détecter des lancements clandestins, des changements dans des sites de production d'armements ou des essais d'armes chimiques et biologiques. Pour contrôler les exportations d'armes, l'IA pourrait améliorer leur traçabilité.

Renforcer la prévention, le maintien et la consolidation de la paix. L'IA permettrait de mener des opérations de maintien de la paix mieux adaptées et donc plus efficace. Le système de traduction instantanée « Resistance » mis en avant par les armées françaises a pour objectif de permettre, hors-connexion et sans réseau, la communication avec la population locale et ainsi de lutter contre la désinformation.

Des risques à atténuer

Des risques propres à la technologie. Les techniques d'apprentissage présentent différents risques de biais : biais involontaires ; biais volontaires ; biais de reconstitution de données particulièrement sensibles ; résultats opaques ou peu explicables. Se pose aussi la question d'une consommation exponentielle de ressources en énergie.

L'aggravation des risques pour la sécurité et la stabilité internationale. Dans de mauvaises mains, l'IA peut aggraver certains risques à la sécurité et la stabilité internationales (scénarios d'escalade, course aux armements, prolifération vers des acteurs non étatiques, extension des opérations d'influence et des actions hostiles dans le domaine cyber) pour lesquels il faudra adapter les mesures d'atténuation des risques. Les risques de déresponsabilisation induits par la dépendance aux technologies nécessitent de garantir la responsabilité humaine.

II. Les principes et mesures clés pour une « intelligence artificielle responsable » tout au long du cycle de vie

Développer une intelligence artificielle respectueuse du droit international humanitaire

Adapter les examens de licéité. Si cet examen s'applique pleinement à l'IA militaire, les modalités précises de cet examen devront s'adapter aux spécificités de cette technologie.

Effectuer des réexamens appropriés. Cet examen doit être conduit, en tant que de besoin, lors des différentes phases du cycle de vie d'un système d'arme. Il doit se tenir lorsqu'un même dispositif subit des innovations ou intègre de nouveaux composants susceptibles de modifier significativement les effets provoqués.

Développer une intelligence artificielle fiable et sécurisée

Évaluer, qualifier et certifier les systèmes. Ces systèmes doivent être évalués et qualifiés, au juste niveau (en fonction de la criticité des fonctions), via une analyse de risque durant la phase de conception. Ils doivent être associés à des cas d'utilisation définis. La révision de ces vérifications doit être envisagée avec une fréquence adaptée aux enjeux.

S'appuyer sur des données maîtrisées et souveraines. Il convient de mettre en place des parades face aux risques d'atteinte aux données, ainsi que des défenses adaptées.

Corriger et réentraîner les systèmes. Il importe de relever et de caractériser les erreurs rencontrées (lors des tests ou de l'usage opérationnel), de sensibiliser les opérateurs au retour d'expérience et de vérifier continuellement si le système est conforme à nos obligations internationales.

Soumettre l'intelligence artificielle à un contrôle humain approprié et une chaîne de commandement responsable

Garantir la conformité de la prise de décision et de l'action avec le droit. L'opérateur ou le chef militaire doit pouvoir exercer son propre discernement pour vérifier si les résultats proposés sont conformes aux ordres donnés et aux obligations juridiques.

Garantir la responsabilité humaine. La responsabilité humaine dans la conception, le déploiement et l'emploi de technologies d'IA constitue un principe indérogeable, nécessitant de formaliser les chaînes de responsabilité de ceux en charges des fonctions de commandement, du contrôle et de l'exécution.

Adapter le contrôle humain. Analyser et caractériser le contrôle humain approprié, sans pour autant brider les capacités du système intégrant des technologies d'IA, est une problématique complexe qui doit prendre en compte différents facteurs humains, techniques et contextuels.

Former les chefs militaires et le personnel pour maitriser ces systèmes. Une phase de formation et d'entraînement doit être instaurée avant l'emploi, afin de sensibiliser le personnel aux apports et aux risques.

Développer une intelligence artificielle durable

Protéger la recherche. L'objet et l'horizon des programmes de recherche doivent être ouverts, sans que soient édictées a priori des interdictions trop larges.

Soumettre la recherche à une réflexion éthique. La France s'est dotée d'une structure de réflexion permanente sur les enjeux éthiques des nouvelles technologies dans le domaine de la défense, à savoir, le Comité d'éthique de la défense.

Développer une intelligence artificielle frugale. Privilégier un comportement frugal implique de réfléchir au recours de l'IA et d'améliorer la résilience et la durabilité des systèmes, tout en contrôlant les coûts.

III. Un processus dédié pour mettre en place une gouvernance globale visant à opérationnaliser les principes d'une intelligence artificielle responsable

Un processus universel et inclusif. Les discussions doivent rassembler toutes les parties prenantes : les États – en particulier, une participation active des États qui développent et emploient les systèmes est absolument nécessaire ; eu égard à cet impératif, le processus décisionnel devra tenir compte des diverses positions et adopter des règles en ce sens en vue d'assurer le consensus – mais aussi le monde industriel, le domaine scientifique et académique et la société civile, de façon à éviter des discussions déconnectées du réel et à préserver l'innovation. La Première Commission de l'Assemblée générale peut constituer une enceinte appropriée.

Une architecture de gouvernance rationalisée et cohérente. Un cadre unique devrait permettre de rationaliser les efforts pour gagner en efficacité et renforcer l'impact des résultats. Il sera indispensable qu'une complémentarité soit assurée avec les discussions du Groupe d'experts gouvernementaux sur les technologies émergentes dans le domaine des systèmes d'armes létaux autonomes, qui doit pouvoir poursuivre ses travaux après 2026, sous un nouveau mandat.

Un processus à visée opérationnelle, centré sur les enjeux propres au domaine militaire. La gouvernance doit s'attacher au corpus juridique applicable aux conflits armés, et donc en premier lieu au droit international humanitaire. La priorité de tout processus international doit être d'assurer le respect des normes juridiques existantes, en discutant de l'établissement de principes directeurs et des moyens de les mettre en œuvre par les États (en facilitant l'échange de bonnes pratiques et en favorisant la coopération et l'assistance internationales, selon des modalités adaptées aux affaires militaires), tout en faisant la promotion de mesures de confiance et de réduction des risques adaptées.

Grèce

[Original : anglais]
[10 avril 2025]

L'intégration de l'intelligence artificielle (IA) dans le secteur de la défense a fondamentalement influencé les moyens et les méthodes de conduite des opérations militaires. Les applications militaires de l'IA ont apporté des avantages opérationnels significatifs, notamment l'amélioration de la vitesse de prise de décision, l'amélioration de la détection et de la prédiction des menaces, la perception et l'évaluation de la situation en temps réel, l'optimisation de l'allocation et de la planification des ressources, le soutien logistique, l'augmentation des capacités humaines face à des tâches complexes et le traitement efficace de données de renseignement à grande échelle.

Cependant, malgré ces avancées, il est essentiel de reconnaître que le progrès technologique soulève également des difficultés complexes et multidimensionnelles qui nécessitent un examen minutieux pour s'assurer que celles-ci ne compromettent pas la paix, la sécurité et la stabilité, tant au niveau régional qu'au niveau mondial.

À cet égard, la Grèce est particulièrement préoccupée par l'utilisation de systèmes militaires dotés de capacités d'apprentissage automatique, qui soulève plusieurs difficultés, notamment en matière de transparence et d'explicabilité, car les modèles complexes peuvent fonctionner comme des « boîtes noires » dont les processus décisionnels ne sont pas définis, en particulier compte tenu de l'évolution constante de l'environnement sur un champ de bataille.

En outre, l'utilisation potentielle de l'IA générative dans le matériel militaire engendre un niveau important de complexité et d'incertitude, car ces systèmes pourraient générer de manière autonome des solutions nouvelles et s'adapter aux conditions changeantes du champ de bataille en analysant et en apprenant continuellement à partir de nouvelles données – des capacités qui sont d'une importance primordiale pour la Grèce. Pour surmonter ces difficultés, il est essentiel d'imposer des limites et des contraintes opérationnelles claires à l'utilisation de ces systèmes, afin d'éviter tout comportement involontaire.

Compte tenu de ce qui précède, l'une des questions les plus préoccupantes concernant l'utilisation de l'IA dans le domaine militaire réside dans son intégration dans les systèmes de commandement, de contrôle et d'aide à la décision liés à l'utilisation des armes nucléaires. La perspective de déléguer les décisions relatives à la dissuasion nucléaire, ou même le lancement des protocoles pertinents relatifs à leur utilisation, à des systèmes dotés d'IA nécessite un examen approfondi afin de garantir à la fois le contrôle et l'implication d'humains dans ces décisions et la mise en place de garde-fous de cybersécurité appropriés pour éviter une escalade involontaire.

Compte tenu de l'environnement géopolitique complexe actuel, il convient également de se pencher sur les efforts consentis par les États pour maintenir leur supériorité militaire, efforts qui pourraient alimenter une course aux armements caractérisée par un manque de transparence et un climat de suspicion réciproque. Cette concurrence peut exacerber l'instabilité géopolitique et mettre sérieusement en péril la sécurité mondiale, dans la mesure où l'équilibre des pouvoirs est rompu et où le fossé technologique entre les États avancés et les États en développement devient de plus en plus prononcé.

En outre, le développement et le déploiement croissants de capacités utilisant l'IA par les forces armées pourraient permettre d'abaisser le seuil des conflits armés. Le rythme accéléré de la prise de décision et la dépendance croissante à l'égard des systèmes de drones sur les théâtres d'opérations augmentent le risque d'escalade involontaire, car l'élément humain est de plus en plus remplacé par des systèmes de drones sur le champ de bataille.

Dans ce contexte, il convient de tenir compte d'un autre paramètre, à savoir la prolifération et le détournement des capacités de l'IA au profit d'États qui ne respectent pas l'ordre international fondé sur des règles et d'acteurs non étatiques, notamment des organisations terroristes. Les technologies de l'IA devenant plus accessibles, il est tout à fait probable que ces acteurs les acquièrent et les déploient pour mener à bien des objectifs déstabilisateurs, ce qui remettrait encore plus en cause la sécurité internationale.

Les applications militaires de l'IA suscitent également des risques et des difficultés liés aux opérations psychologiques et à la mésinformation, car elles favorisent la production massive d'informations fallacieuses, de *deepfakes* et de données falsifiées visant à tromper la population et à déstabiliser les institutions. Les comptes automatisés (robots) et les algorithmes de propagande ciblée renforcent les opérations psychologiques, influencent l'opinion publique et les processus électoraux et créent des tensions sociales, notamment en sapant la confiance des populations dans les opérations de maintien de la paix par des campagnes de désinformation. Les

biais sociaux, tels que ceux liés au genre, à l'âge, à la race et au handicap, suscitent également des inquiétudes, et il est essentiel de mettre en œuvre des évaluations des risques et des mesures d'atténuation pour prévenir les biais et la discrimination involontaires dans les algorithmes.

En outre, les applications de l'IA dans le domaine de la cybersécurité peuvent être utilisées soit pour protéger les infrastructures critiques, soit à des fins malveillantes, telles que les cyberattaques et l'interception de données. Les menaces hybrides combinant des opérations militaires traditionnelles et des tactiques de renseignement offensives imposent une vigilance et une coordination accrues entre les États et les acteurs internationaux afin d'éviter toute escalade et de préserver la paix et la sécurité régionales et internationales.

À la lumière de ce qui précède, la Grèce soutient fermement les efforts internationaux visant à garantir l'utilisation responsable de l'IA dans le domaine militaire, car, malgré les difficultés décrites ci-dessus, cette technologie peut renforcer la mise en œuvre du droit international humanitaire et contribuer à la protection des civils en améliorant la précision des cibles, en renforçant la surveillance et en optimisant l'aide humanitaire.

C'est dans cet esprit que la Grèce a organisé, le 4 avril 2025, avec la France et la République de Corée, et avec le soutien précieux de l'Arménie, de l'Italie et du Royaume des Pays-Bas, une réunion du Conseil de sécurité organisée selon la formule Arria sur le thème de l'intelligence artificielle sûre, inclusive et digne de confiance au service du maintien de la paix et de la sécurité internationales. Cette séance a permis de mieux comprendre comment l'Organisation des Nations Unies pouvait contribuer au maintien de la paix et de la sécurité internationales, notamment par la réglementation, la non-prolifération et la prévention du détournement des capacités d'IA à des fins militaires, le renforcement de l'état de droit, des valeurs démocratiques, de la cohésion sociale et du développement économique.

En outre, dans le cadre de son engagement international, la Grèce a soutenu les déclarations conjointes publiées lors des deux sommets sur l'intelligence artificielle responsable dans le domaine militaire, tenus à La Haye (15 et 16 février 2023) et à Séoul (9 et 10 septembre 2024), sur les mesures à prendre en faveur du développement et de l'utilisation responsables de l'IA dans le domaine militaire. La Grèce a également approuvé la Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles, dirigée par les États-Unis d'Amérique, ainsi que la Déclaration de Paris sur le maintien du contrôle humain dans les systèmes d'armes dotés d'IA.

Par ailleurs, la Grèce a mis en place un comité consultatif de haut niveau[1] sur l'IA chargé d'élaborer une stratégie nationale globale en matière d'IA, ainsi que les structures nécessaires au sein du Ministère de la défense nationale pour surmonter les difficultés technologiques, juridiques, éthiques et politiques découlant des applications de l'intelligence et de l'autonomie artificielles dans le domaine militaire.

Enfin, afin de contribuer de manière constructive au dialogue international sur l'utilisation responsable de l'IA dans le domaine militaire, la Grèce organise une conférence internationale sur le thème des conflits armés et de la gestion de crise à l'ère de l'IA, qui se tiendra à Athènes les 22 et 23 mai 2025.

[1] L'étude historique du Comité, intitulée « Plan d'action pour la transformation de l'IA en Grèce », fournit des principes directeurs et des projets phares visant à stimuler les progrès de l'intelligence artificielle en Grèce, assortis de priorités telles que la sauvegarde et le renforcement de la démocratie, l'atténuation des changements climatiques et l'adaptation à leurs effets, ainsi que le soutien à la sécurité.

Inde

[Original : anglais]
[1er avril 2025]

L'intelligence artificielle (IA) est une technologie porteuse de changement qui affecte considérablement tous les aspects de la vie humaine. Elle est développée à une échelle et à une vitesse sans précédent et est adoptée et déployée rapidement pour toute une série d'applications. L'IA peut avoir des effets transformateurs sur la réduction de la pauvreté et l'amélioration de la vie des populations. Cette réalité est particulièrement pertinente dans le cas de pays en développement comme l'Inde.

Il est nécessaire de faire des efforts collectifs au niveau mondial pour établir une gouvernance et des normes en matière d'IA qui respectent nos valeurs communes, prennent en compte les risques et instaurent la confiance. La gouvernance et les normes en matière d'IA devraient : tenir compte de la profonde interdépendance transfrontalière ; promouvoir l'innovation ; être déployées pour le bien de toutes et tous ; promouvoir l'accès et l'équité afin de faire en sorte que les avantages de l'IA soient accessibles à toutes et tous, en particulier aux pays du Sud. L'Inde s'est engagée à mener des discussions ouvertes sur l'innovation et la gouvernance.

Les discussions sur l'IA dans le domaine militaire doivent être ancrées dans la réalité militaire, où l'on assiste à une intégration rapide de l'IA dans les doctrines et les opérations militaires. Les conflits en cours dans le monde ont démontré à la fois les risques et les possibilités découlant de l'adoption croissante de ces technologies.

Le développement, le déploiement et l'utilisation de l'IA dans le domaine militaire soulèvent des problèmes éthiques, juridiques et de sécurité. Sans minimiser ces difficultés, l'Inde soutient le point de vue qui a été exprimé sur le potentiel de l'IA à améliorer le respect du droit international humanitaire.

L'Inde soutient les efforts collectifs déployés au niveau mondial pour réglementer de manière appropriée le développement, le déploiement et l'utilisation de l'IA dans le domaine militaire. Ces efforts devraient répondre aux préoccupations juridiques et éthiques et permettre de recenser et d'atténuer les risques liés à l'IA dans le domaine militaire.

Tout effort collectif visant à réglementer de manière appropriée l'IA dans le domaine militaire devrait être axé sur les applications et l'utilisation, et non sur la technologie et ses composantes. Il faut éviter de stigmatiser la technologie. L'accès aux technologies à des fins de développement ne doit pas être limité.

L'IA devrait être utilisée légalement dans le domaine militaire, conformément au droit inhérent à la légitime défense individuelle ou collective prévu par le droit international. Le droit international humanitaire continue de s'appliquer pleinement à l'IA dans le domaine militaire. Les principes cardinaux du droit international humanitaire, à savoir la distinction, la proportionnalité et la précaution, s'appliquent à tous les moyens et méthodes de guerre dans le passé, le présent et l'avenir.

Le jugement humain et le contrôle de l'utilisation de l'IA dans le domaine militaire sont essentiels pour atténuer les risques et garantir le respect du droit international humanitaire.

Tout effort collectif ou toute réglementation appropriée concernant l'IA dans le domaine militaire devrait tenir compte des obligations juridiques en vigueur et respecter la juridiction et la compétence nationales, ainsi que les capacités nationales pertinentes.

L'Inde s'est engagée à utiliser l'IA de manière responsable dans le domaine militaire.

L'Inde est en passe d'élaborer un cadre d'évaluation fiable de l'IA dans le secteur de la défense afin de surmonter les difficultés complexes suscitées par les technologies modernes utilisant l'IA. Le cadre s'articule autour de cinq principes clés : a) fiabilité et robustesse ; b) sûreté et sécurité ; c) transparence; d) équité ; e) confidentialité. Ces principes jettent les bases de discussions ultérieures sur la réglementation appropriée du développement, du déploiement et de l'utilisation de l'IA dans le domaine militaire.

Indonésie

[Original : anglais]
[11 avril 2025]

L'Indonésie se félicite de la discussion sur les possibilités et les difficultés que l'application de l'intelligence artificielle dans le domaine militaire présente pour la paix et la sécurité internationales dans le domaine militaire, en mettant un accent particulier sur les domaines autres que les armes létales autonomes, conformément aux paragraphes 7 et 8 de la résolution 79/239 de l'Assemblée générale.

Étant donné que l'IA dans le domaine militaire recouvre un large éventail de systèmes et d'applications, pour organiser une réflexion multilatérale inclusive sur le sujet à l'Organisation des Nations Unies, cette réflexion devrait s'étendre au-delà des capacités cinétiques (telles que les systèmes d'armes létaux autonomes), pour englober les capacités non cinétiques, qui peuvent avoir un objectif hostile (par exemple, les systèmes autonomes de cyberguerre, le brouillage radar adaptatif ou les capacités de guerre électronique) ou de soutien (par exemple, la logistique, l'évacuation sanitaire ou la surveillance tactique) dans le contexte militaire. Elle devrait également couvrir d'autres capacités susceptibles d'avoir un effet direct sur l'équilibre stratégique, telles que l'amélioration de la détection (par exemple par satellite ou anti-sous-marine), du renseignement ou de la planification de la guerre.

L'Indonésie reste fermement attachée au maintien de la paix et de la sécurité internationales, comme le prévoit le préambule de la Constitution indonésienne. Animée par cet engagement, l'Indonésie estime que l'utilisation de l'IA dans le domaine militaire doit être réglementée de manière à promouvoir la paix, la sécurité et les objectifs de développement durable. L'IA doit être un vecteur de paix et de sécurité, et non un facteur d'insécurité, de conflit ou de rivalité stratégique.

Bien que l'IA ne soit pas une arme en soi, l'Indonésie reconnaît qu'elle sert à la fois de multiplicateur de force et d'amplificateur de menace, capable de générer des avantages significatifs et des risques sérieux pour la paix et la sécurité internationales. L'utilisation de l'IA dans le domaine militaire soulève diverses questions éthiques, juridiques, morales et techniques, qui devraient être soigneusement examinées et débattues dans le contexte du respect du droit international, notamment le droit international humanitaire et le droit international des droits humains.

D'une part, on estime que l'IA offre un large éventail de possibilités. Elle peut améliorer le traitement des données ; accroître l'efficacité, la précision et l'exactitude des opérations ; et potentiellement améliorer le respect du droit international humanitaire, notamment en favorisant l'évaluation de la proportionnalité et les mesures de précaution visant à réduire les préjudices causés aux civils. L'IA peut également renforcer les capacités de renseignement, de surveillance et de reconnaissance, appuyer la logistique et la planification et améliorer la gestion du personnel.

D'autre part, l'IA soulève une série de risques et de conséquences, notamment la possibilité d'alimenter les courses aux armements, de proliférer vers des acteurs non étatiques, de favoriser des utilisations criminelles et irresponsables, d'exacerber

le déséquilibre de la puissance militaire grâce à la supériorité technologique et d'accroître l'instabilité, les erreurs de calcul, l'escalade et l'ambiguïté juridique. Au nombre des risques techniques figurent également les cybervulnérabilités, les dysfonctionnements des systèmes, les biais dans les données, les erreurs d'identification des cibles et d'autres incertitudes opérationnelles.

L'Indonésie est particulièrement préoccupée par les risques existentiels découlant de l'intégration potentielle de l'IA dans les systèmes de commandement, de contrôle et de communication nucléaires. L'Indonésie réaffirme sa position de principe selon laquelle l'utilisation et la menace d'utilisation d'armes nucléaires enfreignent le droit international et qu'il convient de prendre des mesures urgentes et décisives pour faire respecter et renforcer les normes contre les armes nucléaires. L'introduction de l'IA dans les systèmes d'armes nucléaires exacerbe les risques existentiels liés à l'utilisation d'armes nucléaires, qu'elle soit intentionnelle, involontaire ou accidentelle, et accroît les dangers nucléaires. Il s'agit d'une menace pour la sécurité de toutes les nations. L'Indonésie demande instamment à tous les États dotés d'armes nucléaires de réévaluer leur dépendance à l'égard de ce type d'armes et de réaffirmer leur engagement collectif en faveur d'un monde exempt d'armes nucléaires. En attendant l'élimination totale des armes nucléaires, les États dotés de telles armes doivent maintenir un contrôle humain significatif, faire montre de responsabilité et respecter leur obligation de rendre des comptes eu égard aux armes nucléaires et à leurs vecteurs dans le contexte du développement de l'IA.

Compte tenu de ces considérations, l'Indonésie préconise une approche prudente pour relever les difficultés liées à l'utilisation de l'IA dans le domaine militaire. L'Indonésie souligne que le développement, l'application et l'utilisation de l'IA dans le domaine militaire doivent être régis afin d'en exploiter les avantages et d'en atténuer les risques. Cette gouvernance doit servir la paix, la sécurité et la prospérité collectives de toutes les nations. En conséquence, l'Indonésie présente les points clés ci-après.

Premièrement, l'Indonésie affirme que le droit international doit être respecté tout au long du cycle de vie des technologies d'IA, notamment la Charte des Nations Unies, le droit international humanitaire, le droit international des droits humains et les traités de désarmement et de non-prolifération. Les États devraient procéder à des examens juridiques à tous les stades, de la passation des marchés à l'évaluation. Les États doivent endosser la responsabilité du développement et de l'application de l'IA dans le domaine militaire, notamment en ce qui concerne la légalité des applications de l'IA dans la conduite de la guerre ou des hostilités. En l'absence de telles lois réglementant l'utilisation de l'IA dans le domaine militaire, il importe de souligner que l'utilisation sera régie par les lois de l'humanité et les impératifs de la conscience publique.

Au-delà du droit international, les considérations éthiques devraient compléter les cadres juridiques pour orienter les modalités de gouvernance de l'utilisation de l'IA dans le domaine militaire. Des principes tels que la traçabilité, l'obligation de rendre compte, la responsabilité, l'explicabilité, l'humanité, la transparence, l'équité et la justice doivent être mis en avant dans le contexte du développement et des applications de l'IA.

Deuxièmement, l'Indonésie souligne le rôle essentiel que joue l'élément humain s'agissant de garantir l'obligation de rendre compte et la responsabilité à tous les niveaux, que ce soit au niveau de l'État, de l'entreprise ou de la personne, dans la conception, le développement, le déploiement et l'utilisation de l'IA dans le domaine militaire.

Le développement, l'application et l'utilisation de l'IA dans le domaine militaire doivent rester centrés sur l'humain et être régis de manière à servir les intérêts de

l'humanité. Le contrôle humain efficace et significatif doit être préservé et renforcé par la formation, en particulier dans les décisions qui prévoient le recours à la force. Les décisions critiques doivent faire appel au jugement, à l'intervention, à la supervision et au contrôle de l'humain. En outre, bien que la notion de « contrôle humain significatif » ait été de plus en plus acceptée dans le contexte de l'encadrement de l'utilisation de l'IA dans le domaine militaire, l'Indonésie estime que ce concept doit encore répondre aux questions juridiques, morales, techniques et réglementaires associées à une telle utilisation. Il faut se mettre d'accord sur ce qu'implique un contrôle humain « significatif » dans la pratique.

Si la gouvernance de l'IA réglementera principalement la conduite des États, elle doit également concerner les parties prenantes civiles, en particulier les entreprises technologiques associées à l'utilisation de l'IA dans le domaine militaire. Les États doivent veiller à ce que le secteur privé respecte le droit international et les normes éthiques tout en soutenant la croissance de l'écosystème de l'innovation en matière d'IA. Il incombe aux chercheurs et aux entreprises de veiller à ce que leurs technologies d'IA soient fiables, sûres, sécurisées, responsables et soumises à un contrôle humain responsable. Ils devraient également être responsables du suivi, de la communication et de la prise en compte des risques liés à leur produit.

Troisièmement, l'Indonésie souligne le besoin urgent de cadres de gouvernance juridique et réglementaire multilatéraux, inclusifs et complets. Ces cadres doivent tenir compte des intérêts de tous les États, quel que soit leur niveau de développement en matière d'IA. Tous les États doivent pouvoir participer sur un pied d'égalité à l'élaboration des règles et des normes régissant l'utilisation de l'IA dans le domaine militaire, afin de garantir une représentation équitable et de favoriser la confiance au niveau mondial.

Une large participation des parties prenantes est essentielle, compte tenu des multiples aspects éthiques, juridiques et techniques de l'IA. L'engagement de diverses disciplines et cultures est également nécessaire pour s'assurer que les systèmes d'IA sont conformes au droit international, au droit humanitaire, aux droits humains et aux engagements en matière de désarmement avant leur application dans la sphère militaire.

Quatrièmement, il est essentiel de rester conscient des risques, des difficultés et des incidences découlant du développement, du déploiement et de l'utilisation de l'IA dans le domaine militaire, qu'ils soient technologiques ou non technologiques, et de favoriser un débat constructif à ce sujet. L'Indonésie souligne la nécessité d'évaluer en permanence les conséquences plus générales de l'utilisation de l'IA à des fins militaires pour la paix et la sécurité internationales, en particulier dans le contexte de la non-prolifération et du désarmement. Des études plus approfondies sont nécessaires pour comprendre ces conséquences, qui ne sont toujours pas suffisamment étudiées.

Le recensement des risques associés au développement, au déploiement et à l'utilisation de l'IA dans le domaine militaire permettra d'établir des prévisions fondées sur des données probantes, d'évaluer les risques et de mettre au point des mesures d'atténuation des risques.

Il est également essentiel de mieux comprendre les risques liés à l'utilisation de l'IA dans le domaine militaire et de sensibiliser les esprits à cette question. À cet égard, il convient de promouvoir la transparence, notamment en communiquant sur les politique et stratégie nationales, en particulier pour recenser, évaluer et atténuer les risques ; en mutualisant les capacités d'IA dans le domaine militaire, le cas échéant, afin d'accroître l'application du principe de responsabilité et les mesures de confiance ; en partageant les enseignements tirés et les meilleures pratiques par-delà les frontières, les industries et les secteurs.

Cinquièmement, la gouvernance de l'IA ne doit pas entraver le développement technologique ni limiter l'accès des pays en développement à l'IA. Les cadres doivent éviter d'imposer des conditions ou des barrières qui restreignent l'accès équitable. Il convient d'adopter une approche équilibrée qui tienne compte des risques tels que la prolifération tout en garantissant l'accessibilité de l'IA aux États disposant de ressources limitées.

Enfin, la gouvernance de l'IA doit mettre l'accent sur la réduction de la fracture numérique et de la fracture de l'IA. Les pays en développement doivent faire face à des contraintes majeures, non seulement en ce qui concerne les capacités d'IA, mais également en ce qui concerne leur aptitude à gérer efficacement ces technologies. Si ce fossé n'est pas comblé, les efforts de gouvernance mondiale seront réduits à néant, car de nombreux États restent mal outillés pour relever les difficultés complexes et transfrontalières que présente l'IA.

L'Indonésie souligne qu'il est urgent de s'attaquer à la fracture numérique et à la fracture de l'IA entre les nations et à l'intérieur de celles-ci, notamment en ce qui concerne l'accès aux ressources financières, humaines et techniques. Ces fractures risquent d'aggraver les inégalités mondiales et d'accroître les risques de conflit.

En tant que biens publics mondiaux, la paix et la sécurité nécessitent une coopération internationale entre tous les pays, tant développés qu'en développement, afin de relever les difficultés communes et de profiter des avantages collectifs, notamment ceux liés au développement, à l'application et à l'utilisation de l'IA dans le domaine militaire. Dans ce contexte, l'Indonésie appelle à une coopération et à une assistance internationales renforcées et équilibrées afin de promouvoir la capacité de l'IA et les cadres de gouvernance au niveau mondial. Cette coopération doit être mise en place sur une base équitable et mutuellement convenue, en tenant compte des besoins et des contextes de chaque pays en développement. Il convient notamment de lancer des initiatives en matière de renforcement des capacités, d'éducation, de transfert de technologie, d'apprentissage tout au long de la vie, de formation technique, de recherche commune et de partage des connaissances.

Cette coopération doit se faire à plusieurs niveaux, non seulement entre les États et les organisations internationales, mais aussi entre les secteurs au sein des pays. Les partenariats public-privé devraient être encouragés afin de promouvoir une innovation responsable et de sensibiliser l'industrie aux incidences que leurs technologies peuvent avoir sur la paix et la sécurité internationales.

La coopération internationale est essentielle non seulement pour réduire la fracture numérique et la fracture de l'IA, mais aussi pour créer un environnement propice à l'instauration d'un climat de confiance entre les États. Elle peut contribuer à réduire les divisions géopolitiques et la concurrence dans le domaine de l'IA. La coopération internationale doit être ancrée dans les principes d'égalité, de confiance, de bénéfice mutuel, de respect de la souveraineté et de solidarité afin d'ouvrir la voie à une collaboration significative, notamment le transfert technologique et le partage des connaissances.

L'Indonésie reconnaît également la valeur du renforcement des mécanismes de coopération régionale qui prennent en compte les particularités locales et régionales. Ces mécanismes peuvent servir de fondements à un consensus mondial plus large, tout en offrant un espace pour des délibérations plus granulaires et adaptées au contexte.

Iran (République islamique d')

[Original : anglais]
[12 mars 2025]

Pour donner suite au paragraphe 7 de la résolution 79/239 de l'Assemblée générale, dans lequel le Secrétaire général a été prié de solliciter les vues des États Membres sur les possibilités et les difficultés que l'application de l'intelligence artificielle (IA) dans le domaine militaire présentait pour la paix et la sécurité internationales, en particulier dans des domaines autres que les systèmes d'armes létaux autonomes, la République islamique d'Iran fait part ci-dessous de ses vues.

L'IA devient l'un des principaux moteurs de changement dans le monde d'aujourd'hui, laissant une marque indélébile sur la manière dont l'industrie de l'armement évoluera dans un avenir proche, ce qui influe profondément sur la paix et la sécurité internationales. Des acteurs étatiques et non étatiques promeuvent activement des ambitions concurrentes dans le domaine de l'IA, et cette situation appelle une réglementation. Compte tenu du rôle prépondérant des acteurs non étatiques et de la nécessité de trouver un équilibre entre les procédures et les tendances en matière de réglementation et d'innovation, il est essentiel que l'autorité réglementaire demeure la prérogative souveraine des États Membres.

Sur le fond, la République islamique d'Iran soutient que, à l'instar d'autres technologies employées dans le cyberespace et l'espace extra-atmosphérique, l'IA doit être exclusivement utilisée de manière pacifique. Elle estime en outre que, si les circonstances s'y prêtent, les entités militaires peuvent elles aussi tirer parti d'une telle utilisation pacifique.

Étant donné que toutes les nations n'ont pas le même niveau de développement, il est primordial de veiller à ce que le fossé numérique ne débouche pas sur une fracture en matière d'IA. Seul le cadre de l'Organisation des Nations Unies, qui repose sur le consensus, peut garantir l'inclusivité de toutes les procédures réglementaires relatives à l'IA. Cette démarche préserve la souveraineté des États Membres, promeut un environnement propice à un développement équitable de l'IA pour tous et offre une flexibilité qui favorise l'innovation et l'essor de cette technologie. Le rôle central joué par l'ONU en ce qui concerne les questions de réglementation liées à l'IA empêche l'adoption d'approches nationales exclusivistes en la matière. L'inclusivité et la recherche du consensus doivent impérativement prévaloir pour cette question qui revêt une importance primordiale.

Malgré les débats en cours sur l'IA dans diverses instances internationales, notre compréhension de la question et de ses répercussions sur la paix et la sécurité internationales demeure lacunaire. Il est encore trop tôt pour dire que le droit international, le droit humanitaire et le droit international des droits humains s'appliquent pleinement à l'IA. Face à l'ampleur de cette situation inédite qui change rapidement, le cadre juridique international devrait sans doute s'adapter et évoluer d'une façon qui lui serait propre.

En ce qui concerne les initiatives internationales de réglementation, la République islamique d'Iran préfère la conclusion d'accords juridiquement contraignants entre les États Membres à l'adoption d'instruments normatifs ou politiques.

Conformément à sa position de principe sur le désarmement, la République islamique d'Iran rejette toute démarche discriminatoire et conditionnelle motivée par des considérations politiques et toute politique de deux poids, deux mesures. Ainsi, la terminologie utilisée par l'Assemblée générale doit refléter l'unité et le consensus. Dans cette optique, les notions telles que « application responsable » sont trop abstraites pour contribuer à la réglementation d'un domaine défini par le concret et la

précision. Une notion aussi abstraite prêterait le flanc à des interprétations erronées et ouvrirait la voie à une démarche politisée. La République islamique d'Iran s'oppose fermement à l'utilisation d'une terminologie aussi subjective. Elle propose de remplacer le terme « application responsable » par « application pacifique » dans tout instrument futur.

Israël

[Original : anglais]
[10 avril 2025]

Israël prend note de l'adoption de la résolution 79/239 de l'Assemblée générale et a l'honneur de soumettre, comme demandé au paragraphe 7 de ladite résolution, sa contribution nationale au rapport que le Secrétaire général présentera à l'Assemblée à sa quatre-vingtième session, dans la perspective de futurs débats entre les États Membres.

Israël estime que le concept d'intelligence artificielle (IA) prête actuellement à plusieurs interprétations, qui peuvent être affinées au fil du temps.

Il est clair que l'utilisation de l'IA dans le domaine militaire devient plus répandue et plus fréquente que jamais. Israël a voté en faveur de la résolution susmentionnée de l'Assemblée générale et encourage les États et les parties prenantes à mener un débat en gardant une attitude professionnelle et apolitique. Cet échange doit prendre en compte les préoccupations légitimes de tous les États, notamment en matière de sécurité, d'aide humanitaire, d'économie et de développement.

Afin de mener un débat sérieux et responsable sur l'IA dans le domaine militaire, susceptible à terme de produire des résultats concrets, nous estimons qu'il faut adopter une démarche pragmatique, équilibrée et progressive.

Étant donné que la technologie ouvre de vastes perspectives dans presque tous les domaines, y compris le domaine militaire, nous nous félicitons de l'examen des retombées potentielles de ces avancées et des moyens de les concrétiser, ainsi que de l'analyse des risques et des solutions permettant de les atténuer. Israël estime que les technologies émergentes, telles que l'IA, peuvent aussi contribuer à promouvoir le respect du droit international humanitaire. Compte tenu de cet atout, elles ne devraient donc pas être stigmatisées.

Israël continue de participer de manière constructive au débat mondial sur l'utilisation de l'IA à des fins militaires. Il a récemment souscrit à la Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles, portée par les États-Unis. Nous nous réjouissons de participer aux prochaines réunions organisées au sujet de la Déclaration, et de continuer de promouvoir une telle utilisation.

Dans le cadre de la Déclaration, ainsi que dans d'autres contextes, des États ont ces dernières années réfléchi à des lignes directrices sur le développement et l'utilisation de l'IA dans le domaine militaire, que ce soit au niveau national ou international. Parmi les principes les plus fondamentaux et les plus largement partagés que l'on retrouve dans ces orientations, et qui peuvent également éclairer le débat relatif à la résolution 79/239 de l'Assemblée générale, figurent les suivants :

- l'utilisation de l'IA à des fins militaires doit être conforme au droit international applicable ;
- elle doit être responsable et renforcer la sécurité internationale ;
- les États doivent veiller à l'application du principe de responsabilité en ce qui concerne l'utilisation des capacités d'IA conformément au droit international

applicable, y compris en exploitant ces capacités dans le cadre d'une chaîne de commandement et de contrôle humains responsables.

Parmi les mesures pratiques que les États devraient prendre pour donner effet à ces principes, on peut citer les suivantes :

- Prendre des mesures, par exemple en procédant à des examens juridiques, pour s'assurer que leurs capacités militaires reposant sur l'IA seront utilisées conformément aux obligations respectives que leur impose le droit international, en particulier le droit international humanitaire ;
- Prendre les mesures voulues pour garantir le développement, le déploiement et l'usage responsables de ces capacités, et les appliquer aux étapes pertinentes du cycle de vie des capacités militaires d'IA ;
- Veiller à ce que le personnel concerné fasse preuve de précaution dans le développement, le déploiement et l'utilisation de ces capacités, y compris les systèmes d'armes qui en sont dotés ;
- Veiller à ce que de hauts fonctionnaires supervisent de manière efficace et convenable le développement et le déploiement des capacités militaires d'IA, destinées à des applications aux conséquences potentiellement graves, notamment les systèmes d'armes dotés de ces capacités ;
- Appuyer des initiatives visant à garantir que les capacités militaires d'IA sont utilisées de manière responsable et légale, et poursuivre des échanges réguliers avec d'autres États sur le déploiement et l'utilisation de ces capacités.

Israël juge utiles les débats multilatéraux inclusifs concernant l'IA dans le domaine militaire et ses répercussions sur la sécurité internationale, car ils pourraient permettre de trouver le juste équilibre entre nécessités militaires et considérations humanitaires.

Italie

[Original : anglais]
[11 avril 2025]

Présidence italienne du Groupe des Sept

L'intelligence artificielle (IA) a été placée au cœur des débats politiques et techniques durant la présidence italienne du G7 en 2024. Lors de leur Sommet qui s'est tenu dans la région des Pouilles, les dirigeants du Groupe des Sept ont reconnu que l'IA avait des répercussions dans le domaine militaire et qu'il était nécessaire d'établir un cadre garantissant son développement et son utilisation responsables.

Du 18 au 20 octobre 2024, la toute première réunion des ministres de la défense du G7 a eu lieu à Naples. À cette occasion, les ministres ont réaffirmé leur détermination à résoudre les problèmes de sécurité de manière cohérente et concrète, à un moment de l'histoire marqué par une grande instabilité. Ils ont en outre souligné qu'il était nécessaire d'adopter une démarche plus concertée en matière de recherche-développement dans le secteur de la défense, y compris en ce qui concerne la mise en commun et la valorisation des connaissances et du savoir-faire, tout en favorisant un environnement sûr pour prévenir tout accès malveillant, afin de maintenir l'avantage concurrentiel, notamment dans le domaine des technologies émergentes et des technologies de rupture.

Enfin, le Groupe des directeurs et directrices des membres du G7 chargés des questions de non-prolifération a reconnu dans une déclaration l'incidence profonde des technologies de rupture émergentes, telles que l'IA, sur la maîtrise des

armements, la non-prolifération et le désarmement, et sur l'avenir des opérations militaires.

I. Intelligence artificielle responsable dans le domaine militaire

L'Italie salue les travaux entamés en 2023 par les Pays-Bas et la République de Corée sur l'intelligence artificielle responsable dans le domaine militaire, qui visent à offrir une plateforme de débat sur les possibilités, les difficultés et les risques majeurs que présentent les applications militaires de l'IA. Lors du deuxième sommet sur l'intelligence artificielle responsable dans le domaine militaire, qui s'est tenu à Séoul en 2024, l'Italie a adhéré au plan d'action (Blueprint for action), un document dans lequel sont décrits les principes clés d'une gouvernance responsable de l'IA, notamment l'importance du respect du droit international, la responsabilité et l'obligation de rendre compte qui incombent à l'être humain, la fiabilité des systèmes d'IA, ainsi qu'une intervention humaine suffisante durant le développement, le déploiement et l'utilisation de l'IA à des fins militaires.

Les États ayant souscrit au plan d'action ont souligné qu'il importait d'empêcher que les technologies d'IA ne soient utilisées pour contribuer à la prolifération des armes de destruction massive, et de veiller à ce qu'elles ne sapent pas les initiatives menées en matière de maîtrise des armements, de désarmement et de non-prolifération. En outre, afin de parvenir à une compréhension commune de la technologie de l'IA et de ses applications dans le domaine militaire, dans le plan d'action, les États sont invités à s'engager à poursuivre les débats, à élaborer des procédures efficaces d'examen juridique et à adopter des mesures de confiance et des mesures adéquates d'atténuation des risques. À cet égard, la mise en commun d'informations et de meilleures pratiques, et la participation active d'autres parties prenantes, sont essentielles pour progresser dans le débat.

II. Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles

L'Italie salue également la Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles. Les États qui ont souscrit à la Déclaration ont dit que l'utilisation de l'IA à des fins militaires peut et doit être éthique et responsable et renforcer la sécurité internationale. Ils ont reconnu qu'un ensemble de mesures devraient être appliquées dans le développement, le déploiement et l'utilisation de l'IA militaire. Les États se sont engagés à réduire au minimum les biais involontaires dans les capacités militaires d'IA, à veiller à ce que la sûreté, la sécurité et l'efficacité de ces capacités fassent l'objet d'essais appropriés et rigoureux, et à instaurer les garde-fous voulus pour déceler et prévenir les conséquences involontaires, et y remédier efficacement le cas échéant. De plus, il est important qu'une chaîne de commandement et de contrôle humain responsable soit définie et que ces capacités militaires d'IA soient utilisées conformément aux obligations internationales.

III. Pacte pour l'avenir

En septembre 2024, les dirigeants du monde ont adopté le Pacte pour l'avenir, dans lequel ils ont réaffirmé leurs engagements internationaux et donné aux États les moyens de gérer les difficultés et possibilités nouvelles et émergentes. Dans la partie consacrée à la Mesure nº 27, les États sont encouragés à tirer parti des possibilités offertes par les technologies émergentes, y compris l'IA, et à remédier aux risques qui y sont associés. Plus précisément, les États Membres continueront d'évaluer les risques liés aux applications militaires de l'IA et les possibilités qu'elles peuvent offrir tout au long de leur cycle de vie, en consultation avec les parties prenantes concernées.

IV. Déclaration de Paris sur le maintien du contrôle humain dans les systèmes d'armes dotés d'IA

L'Italie a récemment souscrit à la Déclaration de Paris sur le maintien du contrôle humain dans les systèmes d'armes dotés d'IA, adoptée en marge du Sommet pour l'action sur l'intelligence artificielle, qui s'est tenu à Paris du 6 au 11 février 2025. Soulignant que la responsabilité et l'obligation de rendre compte ne peuvent jamais être transférées aux machines, les États ayant souscrit à la Déclaration se sont engagés à adopter une approche anthropocentrique dans le développement, le déploiement et l'utilisation des systèmes d'IA dans le domaine militaire. Ils se sont également engagés à veiller à ce que le déploiement de l'IA dans le secteur militaire soit pleinement conforme au droit international et au droit international humanitaire, et à promouvoir la recherche, le développement et l'innovation avec cette technologie.

V. Groupe d'experts gouvernementaux sur les technologies émergentes dans le domaine des systèmes d'armes létaux autonomes

Les avancées rapides en matière d'intelligence artificielle et d'apprentissage automatique ont également des répercussions importantes sur le rôle de l'autonomie dans les systèmes d'armes. L'Italie estime que la Convention sur l'interdiction ou la limitation de l'emploi de certaines armes classiques qui peuvent être considérées comme produisant des effets traumatiques excessifs ou comme frappant sans discrimination, qui est le fruit d'une synergie des compétences diplomatiques, juridiques et militaires des représentants de gouvernements, d'organisations internationales et d'institutions spécialisées, est de loin l'espace le mieux indiqué pour aborder les questions actuelles et émergentes liées au développement et à l'utilisation des systèmes d'armes. Elle contribue activement aux débats menés dans le Groupe d'experts gouvernementaux sur les technologies émergentes dans le domaine des systèmes d'armes létaux autonomes, créé dans le cadre de cette Convention, et s'emploie à faire avancer les échanges sur l'élaboration des éléments d'un futur instrument, conformément au mandat convenu lors de la réunion de 2023 des Hautes Parties contractantes à la Convention.

L'Italie estime que des interdictions et des règles claires devraient être énoncées dans le futur instrument, en vue de son adoption éventuelle comme protocole additionnel à la Convention sur l'interdiction ou la limitation de l'emploi de certaines armes classiques. Suivant cette logique, les systèmes d'armes létaux autonomes qui ne peuvent être mis au point et utilisés conformément au droit international humanitaire seraient *ipso facto* interdits. En revanche, les systèmes dotés d'une autonomie de décision dans leurs fonctions critiques, qui peuvent être mis au point et utilisés en pleine conformité avec le droit international humanitaire, seraient réglementés. Selon l'Italie, l'élément humain est en effet déterminant à toutes les étapes du cycle de vie des systèmes d'armes létaux autonomes, c'est-à-dire pendant la conception, le développement, la production, le déploiement et l'utilisation. Un niveau suffisant de discernement et de contrôle humain doit être maintenu afin de garantir la responsabilité et l'obligation de rendre compte qui découlent du droit international humanitaire.

Japon

[Original : anglais]
[11 avril 2025]

Dans sa résolution 79/239, l'Assemblée générale a prié le Secrétaire général de solliciter les vues des États Membres et des États observateurs sur les possibilités et les difficultés que l'application de l'intelligence artificielle (IA) dans le domaine

militaire présentait pour la paix et la sécurité internationales, en particulier dans des domaines autres que les systèmes d'armes létaux autonomes, et de lui présenter, à sa quatre-vingtième session, un rapport de fond résumant ces vues et répertoriant les propositions normatives existantes et nouvelles, assorti d'une annexe contenant ces vues, dans la perspective de futurs débats entre les États. Le Japon communique ci-dessous ses vues sur ce sujet afin de contribuer à l'élaboration du rapport et de faire avancer le débat sur la question.

I. Vues générales

Le Japon s'est engagé à préserver et à renforcer un ordre international libre et ouvert, fondé sur l'état de droit, afin que tous les peuples puissent jouir de la paix, de la stabilité et de la prospérité, et à promouvoir la diplomatie pour instaurer un monde sûr et sécurisé dans lequel la dignité humaine est protégée. Dans cet esprit, il est activement investi dans des initiatives visant à renforcer la paix et la sécurité internationales ainsi que la maîtrise des armements et le désarmement.

Le Japon estime que l'application de l'IA dans le domaine militaire doit être examinée de manière approfondie, sur la base d'une bonne compréhension de ses risques et de ses avantages et en tenant compte à la fois des considérations humanitaires et des questions de sécurité. Il est utile de mieux comprendre cette application et de promouvoir des initiatives réalistes et pratiques permettant de veiller à ce qu'elle se fasse de manière responsable, afin d'en maximiser les avantages et d'en atténuer les risques.

En outre, concernant l'application de l'IA dans le domaine militaire, le Japon souscrit à l'idée que, premièrement, le droit international existant s'applique aux questions qu'il régit et qui se posent tout au long du cycle de vie des technologies de l'IA ; deuxièmement, les capacités reposant sur l'IA doivent être utilisées de manière responsable; troisièmement, les êtres humains restent responsables de leur utilisation et de leurs effets et doivent en répondre. Le Japon insiste également sur la nécessité de renforcer la transparence, qu'il considère comme une mesure de confiance importante pour optimiser les avantages et réduire les risques.

II. Vues et approche du Japon concernant les possibilités et les difficultés que l'application de l'intelligence artificielle dans le domaine militaire présente pour la paix et la sécurité internationales

Possibilités

Vues

Les progrès rapides de la science et de la technologie, notamment l'intelligence artificielle, modifient fondamentalement le paradigme de la sécurité. Les pays s'efforcent de développer des technologies de pointe susceptibles de transformer radicalement la nature de la guerre et donc de changer la donne, et il est devenu extrêmement difficile dans la pratique de faire la distinction entre les technologies à usage civil et celles conçues à des fins de sécurité. L'IA recèle un potentiel extraordinaire qui peut transformer tous les aspects des affaires militaires, par exemple : les opérations, le commandement et le contrôle, le renseignement, la surveillance et la reconnaissance, la formation, la gestion de l'information et le soutien logistique. Compte tenu de ses diverses applications dans le domaine militaire, elle peut permettre d'améliorer la précision, l'exactitude et l'efficacité, de renforcer la compréhension et la perception de la situation, de faciliter l'analyse rapide de l'information, de réduire les erreurs humaines et d'économiser la main-d'œuvre. Utilisée à bon escient, l'IA peut contribuer à mieux protéger les civils en situation de conflit et durant la consolidation de la paix après les conflits.

Approche du Japon concernant l'exploitation des « possibilités »

Dans le cadre de l'application de l'IA dans le domaine militaire, il y a lieu de déterminer si une telle utilisation permet de résoudre les problèmes relevés par les humains, sans perdre de vue les capacités et les limites de l'IA. L'application de l'IA ne devrait pas être une fin en soi ni être envisagée indépendamment de ces capacités et limites. Par conséquent, il faut que les États veillent à ce que les capacités militaires reposant sur l'IA aient des utilisations précises et bien définies et qu'elles soient conçues et développées en conséquence. Dans cette optique, il importe de promouvoir une compréhension internationale commune de l'IA, de ses atouts et limites dans le domaine militaire, ainsi que de ses applications potentielles dans ce secteur. En ce qui concerne l'application de l'IA par les autorités chargées de la défense, le Ministère japonais de la défense a publié en juillet 2024 sa Politique générale sur la promotion de l'utilisation de l'IA, dans laquelle il a fait le point sur ses réflexions quant à ces atouts et limites, ainsi que les domaines d'application qu'il privilégiait. Dans ladite politique, à la lumière des capacités et des limites actuelles de l'IA, il a recensé les sept axes prioritaires ci-dessous en matière d'application de l'IA :

- détection et reconnaissance des cibles ;
- collecte et analyse de renseignements ;
- commandement et contrôle ;
- opérations de soutien logistique ;
- systèmes sans équipage ;
- cybersécurité ;
- amélioration de l'efficacité dans l'exécution de tâches administratives.

Dans la Politique générale susmentionnée, le Ministère de la défense indique également qu'il importe de garder à l'esprit que l'IA sert à appuyer la prise de décision humaine et que l'intervention humaine est essentielle lors de son utilisation.

Difficultés

Vues

L'application de l'IA dans le domaine militaire peut présenter des risques d'utilisation abusive ou malveillante, d'escalade et d'abaissement du seuil de déclenchement de conflit, qui peuvent provenir de biais, de conséquences involontaires et d'autres facteurs. À cet égard, le Japon souligne la nécessité d'empêcher que l'IA ne soit utilisée pour contribuer à la prolifération des armes de destruction massive par des États et des agents non étatiques, et insiste sur le fait que l'IA devrait promouvoir, et non entraver, les initiatives de désarmement, de maîtrise des armements et de non-prolifération.

L'approche du Japon pour la résolution des difficultés

Compte tenu des risques que représentent notamment les biais et les utilisations abusives ou malveillantes, le Ministère japonais de la défense s'efforcera de réduire les risques liés à l'IA en s'inspirant du concept d'une IA centrée sur l'humain et des principes de sûreté, d'équité, de protection de la vie privée, de sécurité, de transparence et de responsabilité, comme énoncés dans les orientations sur l'IA publiées en avril 2024 à l'intention des entreprises japonaises. Il s'intéresse également aux débats en cours au sein de la communauté internationale et avec les autorités de défense d'autres pays.

En outre, le Japon suit de près les effets que pourraient avoir les technologies émergentes, telles que l'IA, sur le désarmement et la non-prolifération nucléaires. À cet égard, il se félicite de l'engagement pris en 2022 par les États-Unis, le Royaume-Uni et la France à la Conférence des Parties chargée d'examiner le Traité sur la non-prolifération des armes nucléaires de maintenir le contrôle humain et l'intervention humaine à toutes les étapes essentielles à la formation de décisions souveraines concernant l'emploi des armes nucléaires et à leur exécution. Il invite les autres États dotés d'armes nucléaires à faire de même. En outre, dans une recommandation faite à la Conférence d'examen de 2026, le Groupe international de personnalités éminentes pour un monde exempt d'armes nucléaires a souligné la nécessité de coopérer pour remédier aux problèmes liés aux technologies émergentes et tirer parti des possibilités offertes.

III. Vues sur l'avenir des débats et de la coopération internationale

Il faut adopter une démarche souple, équilibrée et réaliste en matière de gouvernance de l'IA dans le domaine militaire, afin de suivre le rythme du développement et de l'évolution rapides des technologies. Le Japon souligne que les initiatives visant à promouvoir une utilisation responsable de l'IA à des fins militaires peuvent être menées parallèlement aux activités de recherche, de développement, d'expérimentation et d'innovation liées à l'IA, sans les entraver.

Il convient de noter que les débats qui remettent en question certaines technologies d'IA peuvent freiner le développement et l'innovation technologiques dans le secteur civil, et produire un effet dissuasif. En outre, l'application de l'IA dans le domaine militaire devrait faire l'objet d'un débat ouvert associant toutes les parties prenantes.

Compte tenu de ce qui précède, le Japon appuie fermement les résultats des sommets sur l'intelligence artificielle responsable dans le domaine militaire et la Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles, et espère que d'autres États se joindront à ces initiatives.

En ce qui concerne les systèmes d'armes létaux autonomes, il convient de noter que le Japon soutient fermement la poursuite des débats dans le cadre de la Convention sur l'interdiction ou la limitation de l'emploi de certaines armes classiques qui peuvent être considérées comme produisant des effets traumatiques excessifs ou comme frappant sans discrimination. Il compte que les débats sur l'application de l'IA dans le domaine militaire viendront compléter et renforcer ceux menés par le Groupe d'experts gouvernementaux sur les technologies émergentes dans le domaine des systèmes d'armes létaux autonomes, créé en vertu de la Convention.

Le Japon reconnaît qu'il importe de faire preuve de transparence dans l'application de l'IA dans le domaine militaire, car il s'agit d'une mesure de confiance qui contribue à l'atténuation des risques et favorise une collaboration et une coopération efficaces entre les pays. Il reconnaît également qu'il est essentiel de renforcer les capacités pour promouvoir une démarche responsable dans le développement, le déploiement et l'utilisation de l'IA à des fins militaires, et s'engage à intensifier la coopération internationale en matière de renforcement des capacités en vue de remédier au manque de connaissances concernant une telle approche. À cet égard, des méthodes telles que la mise en commun de bonnes pratiques et des enseignements à retenir seront utiles, et le Japon saisira les occasions d'échanger des vues avec d'autres pays.

Enfin, en ce qui concerne l'application de l'IA dans le domaine militaire, le Japon continuera à participer de manière active et constructive aux débats menés au niveau international dans le but de parvenir à une compréhension commune au sein

de la communauté internationale au moyen de débats équilibrés qui prennent en compte les considérations humanitaires et les questions de sécurité.

Lituanie

[Original : anglais]
[9 avril 2025]

La Lituanie se félicite de l'occasion qui lui est donnée de contribuer au rapport du Secrétaire général, comme suite à la résolution 79/239 de l'Assemblée générale. Elle a eu le plaisir d'appuyer cette résolution, adoptée le 24 décembre 2024.

La Lituanie note que le développement et l'utilisation de l'intelligence artificielle (IA) dans le domaine militaire présentent à la fois des possibilités et des difficultés pour la paix et la sécurité internationales. Elle accorde une grande importance à l'élaboration de normes et de principes visant à promouvoir une utilisation responsable, qui permettraient aux États de tirer parti de l'IA dans le domaine militaire et d'en atténuer les risques. Elle croit fermement que tous les États responsables ont intérêt à garantir l'application responsable de l'IA dans ce domaine. Elle est convaincue que, pour faire face aux conséquences de l'utilisation de l'IA à des fins militaires, il faut une action mondiale et une approche multipartite mobilisant les secteurs public et privé, la société civile et le monde universitaire.

La Lituanie soutient fortement la Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles, à laquelle elle a adhéré le 13 novembre 2023. Cette Déclaration contient des principes non juridiquement contraignants et de meilleures pratiques visant à garantir une utilisation responsable et légale de l'IA dans le contexte militaire. On y prend en compte des mesures telles que l'examen juridique, le contrôle approprié, la réduction des biais involontaires et la garantie que les capacités d'IA destinées au domaine militaire ont des cas d'utilisation explicites et bien définis. La Lituanie encourage vivement davantage d'États à signer la Déclaration.

En outre, la Lituanie souscrit à la stratégie de l'Organisation du Traité de l'Atlantique Nord pour l'IA, adoptée en 2021 et révisée en 2024. Dans cette stratégie, l'Organisation définit six principes pour une utilisation responsable de l'IA dans le domaine militaire, à savoir : la légalité, la responsabilité, l'intelligibilité et la traçabilité, la fiabilité, la gouvernabilité et l'atténuation des biais. Ces principes non juridiquement contraignants sont censés concerner l'ensemble du cycle de vie d'une application de l'IA, et la Lituanie s'est engagée à les respecter.

Enfin, la Lituanie se réjouit d'expliciter ses vues sur les possibilités et les difficultés que l'application de l'IA dans le domaine militaire pose à la sécurité internationale. Elle estime que l'utilisation de l'IA à des fins militaires peut et doit se faire de manière responsable avant tout pour renforcer la sécurité nationale de l'État, et contribuer à l'application du droit international, y compris le droit international humanitaire, et au respect des diverses obligations qui incombent à l'État en matière de protection des civils. Outre le renforcement de la protection des civils en période de conflit armé, l'IA responsable offre des possibilités d'améliorer la prise de décision, la logistique, la planification et d'autres opérations visant à accroître l'efficacité.

S'agissant des risques liés à l'IA dans le domaine militaire, la Lituanie souligne les défis qui incluent, sans s'y limiter, la cybersécurité, les biais involontaires dans les capacités d'IA militaire et le comportement inattendu des systèmes reposant sur l'IA. Elle estime que la meilleure façon de gérer ces risques est d'appliquer des principes d'utilisation responsable, de renforcer les capacités et de bien former le personnel aux cas d'utilisation de l'IA et aux systèmes reposant sur cette technologie.

Elle souligne que, afin de tirer parti des avantages de l'IA dans le domaine militaire et de l'utiliser comme une capacité de défense essentielle, les États doivent éviter d'imposer des restrictions excessives et inutiles qui entravent l'innovation en la matière, notamment parce que certains États irresponsables pourraient refuser d'accepter de telles contraintes sur l'IA militaire.

Mexique

[Original : espagnol]
[10 avril 2025]

L'intelligence artificielle, les systèmes d'armes autonomes et le défi mondial que représente leur réglementation

Le Mexique présente ce document comme suite à la résolution 79/239 de l'Assemblée générale intitulée « L'intelligence artificielle dans le domaine militaire et ses conséquences pour la paix et la sécurité internationales ».

Le Mexique reconnaît que l'utilisation de l'intelligence artificielle (IA) dans le domaine militaire peut présenter des avantages. Cependant, elle crée également des défis majeurs en matière de paix et de sécurité internationales, que la communauté internationale doit traiter de façon urgente et coordonnée.

Le Mexique apprécie les échanges multilatéraux promus dans le cadre de l'Organisation des Nations Unies, tels que le premier webinaire de la série « MAPS Dialogue on Military AI: Opportunities, Risks, and International Peace & Security », organisé par le Bureau des affaires de désarmement de l'Organisation, qui contribuent à générer une compréhension commune des risques émergents et des responsabilités partagées. Nous convenons que l'intégration de l'IA dans les applications militaires pose des problèmes fondamentaux pour la paix et la sécurité internationales, notamment l'escalade involontaire des conflits, l'ambiguïté stratégique et l'autonomie croissante dans l'emploi de la force.

Le Mexique considère que la priorité est de consolider la coopération internationale, de promouvoir la transparence, de partager les meilleures pratiques, de renforcer les capacités en vue d'une culture de conformité réglementaire et de respect du droit international, et de progresser dans l'élaboration de cadres réglementaires garantissant l'application de principes éthiques, juridiques et humanitaires dans le développement et le déploiement de l'IA dans des contextes militaires, l'objectif étant d'empêcher cette technologie d'aggraver les asymétries ou d'éroder la stabilité internationale.

Paix et sécurité internationales

Le Mexique doit prendre des mesures pour empêcher la prolifération et l'utilisation abusive de ces technologies, y compris si elle est le fait d'acteurs non étatiques ou si elle a lieu en dehors de cadres juridiques clairs.

La paix et la sécurité internationales ne peuvent être subordonnées à l'intégration de technologies nouvelles et émergentes dans le domaine militaire. Cette intégration doit répondre à une perspective de développement humain et d'avancement social, en particulier en faveur des pays en développement. Ces technologies devraient donc être utilisées à des fins de paix et de règlement des différends, plutôt que pour améliorer l'efficacité du matériel militaire.

La sophistication croissante des menaces numériques et le potentiel d'utilisation des technologies émergentes comme vecteurs d'attaques d'État à État, ainsi que la difficulté de garantir la fiabilité et la précision des systèmes autonomes dans des

contextes militaires, l'exposition aux vulnérabilités dans le cycle de vie de l'IA, les biais algorithmiques, l'empoisonnement des données et l'utilisation de modèles génératifs à des fins malveillantes, soulignent que les risques doivent être atténués de façon proactive.

Les progrès scientifiques et technologiques, en particulier dans les domaines de l'IA, des systèmes autonomes, des technologies numériques et quantiques, dépassent la capacité actuelle des cadres réglementaires de gérer ces risques. Le Mexique réaffirme donc qu'il faut élaborer des cadres de gouvernance globaux, favoriser la coopération internationale et le dialogue multilatéral, et accorder la priorité à la transparence, au principe de responsabilité et à un contrôle humain important tout au long du cycle de vie de ces technologies, sans oublier des essais rigoureux et des garanties éthiques pour leur déploiement.

En l'absence de cadres juridiques internationaux clairs et du consensus multilatéral nécessaire, l'utilisation du terme « responsable » dans ce contexte ne doit pas être interprétée comme une validation ou une acceptation tacite de l'utilisation ou du développement de capacités militaires autonomes basées sur l'IA. Le principe de responsabilité doit nécessairement être lié à la légalité et à l'obligation de rendre des comptes.

Selon cette logique, il est essentiel que le Mexique mette en place des mécanismes de gouvernance et de réglementation qui réduisent la probabilité que l'IA et d'autres technologies de rupture soient utilisées à des fins hostiles, compte tenu du fait que les risques ne se limitent pas seulement à leur déploiement opérationnel mais apparaissent dès les premières étapes de leur conception et de leur développement.

Contextes opérationnels

Compte tenu des différents contextes opérationnels du domaine militaire dans lesquels cette technologie pourrait être incorporée, le Mexique note que l'IA peut avoir des effets distincts.

En cas de conflit armé, il est impératif de veiller à ce que toute technologie basée sur l'IA soit utilisée dans le respect du droit international humanitaire, en particulier des principes de distinction, de proportionnalité, de précaution et d'humanité.

Dans le domaine des opérations de maintien de la paix et des interventions en cas de catastrophe, l'IA peut apporter une contribution positive à la coordination logistique, à la prévision des risques et à la prise en charge des populations touchées, à condition que les droits humains soient pleinement respectés.

En ce qui concerne la sécurité des frontières, le Mexique reconnaît que l'IA peut renforcer les capacités de surveillance. Toutefois, il souligne qu'il importe de garantir le respect de la dignité de toutes les personnes, en évitant les décisions automatiques qui perpétuent les pratiques discriminatoires.

Systèmes d'armes létaux autonomes

Le Mexique considère que les systèmes d'armes létaux autonomes, qui constituent une source de préoccupation particulière pour la paix et la sécurité internationales, sont un élément fondamental de cette discussion. À cet égard, nous préconisons de ne pas fragmenter les débats multilatéraux sur l'intégration des nouvelles technologies dans le domaine militaire et pensons que les systèmes d'armes létaux autonomes devraient être pleinement pris en compte dans ces échanges.

Le Mexique considère qu'il est urgent que la communauté internationale mette en place des interdictions et des réglementations claires sur les systèmes d'armes

létaux autonomes en raison de leur incompatibilité avec le droit international humanitaire et des risques éthiques, juridiques et sécuritaires qu'ils présentent.

Le Mexique a promu les résolutions 78/241 et 79/62 de l'Assemblée générale sur les systèmes d'armes létaux autonomes et s'est porté coauteur de ces textes afin de consolider un espace multilatéral légitime pour relever ces défis.

Le Mexique soutient l'appel lancé par le Secrétaire général et le Comité international de la Croix-Rouge demandant l'ouverture de négociations sur un instrument juridiquement contraignant établissant les interdictions et réglementations nécessaires sur les systèmes d'armes létaux autonomes d'ici à 2026, comme le prévoit le Nouvel Agenda pour la paix.

Le Mexique a manifesté son engagement politique sur cette question en participant à la Conférence de San José (2023), en adhérant au Communiqué de Belém et en participant activement à la conférence internationale sur la question de la réglementation des systèmes d'armes autonomes intitulée « Humanity at the crossroads: autonomous weapons systems and the challenge of regulation » (Vienne, 2024), et en a approuvé le rapport final.

Le Mexique estime que les systèmes d'armes létaux autonomes présentent de multiples risques, notamment :

- L'exclusion du jugement humain dans les décisions critiques d'emploi de la force,
- Le remplacement de l'indispensable évaluation contextuelle dans les opérations militaires,
- L'affaiblissement des dispositifs d'application du principe de responsabilité et d'attribution des responsabilités.

La responsabilité de l'emploi de la force ne doit jamais être transférée à une machine. Les décisions relatives au déploiement, à l'activation ou à la désactivation des systèmes armés doivent à tout moment être prises par des êtres humains soumis à la responsabilité juridique.

Le Mexique réaffirme que toute technologie militaire, y compris celle qui utilise l'intelligence artificielle, doit respecter les obligations internationales découlant de :

- la Charte des Nations Unies ;
- le droit international humanitaire ;
- le droit international des droits humains ;
- le droit pénal international ;
- le droit de la responsabilité internationale.

Dans ce contexte, le Mexique considère qu'il est indispensable d'interdire les systèmes d'armes dont la technologie :

- empêche de faire la distinction entre les cibles militaires et civiles ;
- rend impossible l'application du principe de proportionnalité en lien avec les dommages collatéraux ;
- ne prévoit pas de mécanismes d'annulation s'il s'avère que l'attaque ne doit pas être menée ;
- cause des souffrances inutiles ou des blessures superflues.

Le Mexique insiste sur le fait qu'il est urgent d'entamer des négociations relatives à un instrument juridiquement contraignant qui établisse des interdictions et

des réglementations précises et applicables aux systèmes d'armes létaux autonomes, fasse en sorte qu'un contrôle humain marqué soit maintenu dans les activités critiques, et comprenne des mécanismes efficaces d'application, de suivi et de responsabilisation.

Avantages et risques

En ce qui concerne les utilisations particulières de l'IA, le Mexique dénombre à la fois des avantages et des risques dans les domaines suivants :

- Commandement et contrôle : dans certaines conditions, l'IA pourrait améliorer l'efficacité des décisions opérationnelles, mais celles-ci doivent faire l'objet d'un contrôle humain important, en particulier lorsqu'il s'agit de l'emploi de la force. On sait que l'IA a la capacité de traiter et d'analyser de grands volumes de données et d'informations, dépassant de loin les capacités humaines, ce qui permet de prévoir plus rapidement, plus facilement et plus efficacement les tendances futures et d'éclairer les décisions stratégiques en temps réel.
- Opérations cyber : l'IA offre des capacités précieuses pour ce qui concerne la prévention des problèmes liés au domaine cyber et leur résolution, mais elle accroît également les risques d'escalade des tensions, y compris celui d'une utilisation hostile automatisée sans surveillance adéquate.
- Gestion de l'information et logistique : le traitement massif des données par l'IA peut faciliter les décisions en temps réel, mais il doit s'accompagner de protocoles garantissant une utilisation éthique, explicable et responsable.

En dépit de ce qui précède, le Mexique souligne les risques technologiques associés à l'intégration de l'IA dans des contextes militaires, étant donné que les faits suggèrent la persistance de défaillances techniques ou d'erreurs imprévues susceptibles d'aggraver un conflit.

Norvège

[Original : anglais]
[11 avril 2025]

La Norvège se félicite de l'occasion qui lui est donnée de présenter ses vues sur les possibilités et les difficultés que l'application de l'intelligence artificielle (IA) dans le domaine militaire présente pour la paix et la sécurité internationales, en particulier dans des domaines autres que les systèmes d'armes létaux autonomes, conformément à la résolution 79/239 de l'Assemblée générale, intitulée « L'intelligence artificielle dans le domaine militaire et ses conséquences pour la paix et la sécurité internationales ».

Comme le reconnaît le Secrétaire général dans sa note d'orientation de juillet 2023 intitulée « Un Nouvel Agenda pour la paix », l'IA est à la fois une technologie habilitante et une technologie de rupture qui se voit intégrée à un nombre toujours croissant d'applications civiles, militaires ou à double usage. De plus en plus présente, facilement extensible, peu transparente et évoluant très rapidement, cette technologie pourrait faire peser des risques sur la paix et la sécurité internationales et pose des problèmes de gouvernance.

En tant que fervente défenseure du droit international, du multilatéralisme et de l'innovation responsable dans le secteur de la défense, la Norvège appuie les initiatives visant à promouvoir une communauté des vues, à renforcer la gouvernance et à élaborer une réglementation adéquate de l'IA dans le domaine militaire. À tout le moins, les applications d'IA dans le domaine militaire doivent être développées, déployées et utilisées de manière responsable tout au long de leur cycle de vie, dans

le respect du droit international applicable, en particulier du droit international humanitaire.

Il est important de noter que, dans sa résolution 79/239, l'Assemblée générale a affirmé que le droit international, notamment la Charte des Nations Unies, le droit international humanitaire et le droit international des droits humains, s'appliquait à l'utilisation de l'IA dans le domaine militaire, et souligné qu'il importait de veiller à l'application responsable et centrée sur l'être humain de l'IA dans ce domaine.

En tant que technologie habilitante, l'IA recèle un potentiel extraordinaire pour transformer tous les aspects des affaires militaires, notamment les acquisitions, le matériel, les logiciels, les opérations, le commandement et le contrôle, les communications stratégiques, la surveillance, le renseignement, la formation, la gestion de l'information et le soutien logistique. L'application de l'IA dans le domaine militaire présente des possibilités et des risques prévisibles et imprévisibles, tant au niveau tactique que stratégique. En tant que technologie polyvalente, l'IA est un multiplicateur de force capable de redéfinir la conduite de la guerre. La convergence technologique entre l'intelligence artificielle, la neurotechnologie, la biologie synthétique et l'informatique quantique ajoute encore à la complexité.

Il est fondamental que le développement, le déploiement, l'utilisation et la gouvernance de l'IA se fassent de manière responsable, conformément aux principes éthiques fondamentaux, dans le strict respect des obligations imposées aux États par le droit international, notamment le droit international humanitaire et le droit international des droits humains, et en plaçant l'identification et l'atténuation des risques au centre des activités.

La Stratégie norvégienne pour l'intelligence artificielle dans le secteur de la défense (2023) met en lumière les domaines clés, autres que les systèmes d'armes létaux autonomes, dans lesquels l'IA peut être un atout :

- **Amélioration de la perception de la situation et de l'aide à la décision**. L'utilisation de l'IA dans les domaines du renseignement, de la surveillance et de la reconnaissance est non seulement possible, mais aussi nécessaire, étant donné que les volumes de données importants et croissants ne peuvent être analysés manuellement. L'IA peut servir à filtrer les données, par exemple en les prétraitant, à faire de la traduction automatique, à détecter des objets précis dans des images, à repérer des anomalies et des répétitions, ainsi qu'à recouper les informations pour déceler les tentatives de désinformation. Les améliorations dans ce domaine peuvent conduire à des opérations plus efficaces et plus précises et à une réduction des pertes en vies humaines.
- **Cyberdéfense**. La transition numérique et la dépendance accrue à l'égard des technologies de l'information et des communications s'accompagnent de vulnérabilités, en plus des avantages qu'elles offrent. L'espace numérique offre aux acteurs malveillants la possibilité de commettre des violations de données, de se livrer à l'espionnage et au sabotage et de mener des campagnes d'influence. L'IA peut renforcer la capacité du secteur de la défense à détecter, surveiller, signaler, gérer et contrer les menaces numériques. Elle peut notamment permettre de dresser un tableau plus complet des objectifs et des relations complexes, de recueillir des informations provenant de sources pertinentes et de rationaliser l'utilisation des analyses. Il est essentiel de renforcer les connaissances et compétences sur la manière dont l'IA peut constituer une menace numérique afin de pouvoir détecter et prévenir les attaques numériques à l'avenir. Il est par conséquent crucial que l'IA soit un élément clé du renforcement de la défense du secteur face aux menaces numériques, au moyen d'instruments existants et futurs.

- **Logistique**. La réussite et l'efficacité des opérations militaires reposent sur un soutien logistique efficace. En rationalisant la logistique à l'aide de systèmes intégrant l'IA, on peut renforcer la capacité opérationnelle et améliorer l'état de préparation. L'utilisation de l'IA est déjà bien poussée dans la logistique civile, et plusieurs de ses applications pourraient facilement être adaptées au secteur militaire.
- **Activités d'appui**. De nombreuses activités d'appui militaire pourraient être améliorées et rationalisées à l'aide de l'IA. Il s'agit notamment de tâches visant à appuyer et à renforcer la capacité opérationnelle, telles que l'exploitation et l'entretien du matériel, l'acquisition, la gestion et la liquidation du matériel et des bâtiments, le recrutement, la formation et la gestion du personnel, ainsi que la fourniture de services communs, tels que la comptabilité et l'archivage. L'IA a le potentiel de renforcer les activités d'appui en permettant d'optimiser l'exploitation des données à des fins d'analyse et d'aide à la décision, d'automatiser les tâches et d'améliorer la capacité de gestion de l'information et des connaissances. Elle peut ainsi permettre de passer à un modèle de maintenance prédictive, d'améliorer le flux d'informations, d'instaurer ou de renforcer les systèmes d'appui à la gestion des ressources humaines, et d'affiner la modélisation de l'évolution des coûts liés au matériel et aux bâtiments. L'intégration réussie de l'IA dans les activités d'appui pourrait donc permettre de réduire le temps consacré à ces activités et d'accroître l'efficacité.

En outre, les applications de l'IA dans le domaine militaire peuvent améliorer la mise en œuvre du droit international humanitaire et contribuer aux initiatives visant à protéger les civils et les biens de caractère civil dans les conflits armés. Elles peuvent également être utiles pour les activités de consolidation et de maintien de la paix, et aider à renforcer les capacités de vérification et de contrôle de la maîtrise des armements, du désarmement et d'autres régimes de conformité.

Par ailleurs, l'utilisation de l'IA dans le domaine militaire présente des difficultés sans précédent. L'IA comporte des vulnérabilités inhérentes qui peuvent avoir des conséquences involontaires et conduire à la dégradation d'un contrôle humain réel ainsi que de la responsabilité et de l'obligation de rendre compte qui incombent à l'être humain. L'utilisation de l'apprentissage profond est susceptible de rendre le fonctionnement des modèles d'IA difficile à comprendre, à expliquer et à prédire. Le manque de compréhension peut, par exemple, rendre la dynamique d'escalade des conflits plus opaque et imprévisible.

Des garanties efficaces doivent être mises en place pour veiller à ce que les êtres humains conservent une supervision et un contrôle réels sur le développement, le déploiement et l'utilisation de l'IA. Ce point revêt une importance particulière à mesure que l'application se rapproche des opérations de combat et de l'emploi de la force, notamment dans le cas des systèmes informatisés d'aide à la décision. L'obligation de rendre compte et la responsabilité liées à l'utilisation de l'IA militaire et à ses conséquences doivent toujours incomber aux êtres humains.

Les systèmes d'IA peuvent être très sensibles à la qualité et à la représentativité des données d'entraînement. Des biais, la désinformation et la mésinformation, ou des données d'entraînement incomplètes peuvent donner lieu à des modèles qui produisent des résultats inexacts ou discriminatoires. Le biais d'automatisation peut entraîner une dépendance excessive de l'utilisateur humain aux résultats générés par le système.

Dans le cyberdomaine, les capacités de réponse hautement automatisées ou autonomes, notamment celles qui ne font pas assez intervenir des êtres humains, pourraient provoquer des conséquences involontaires et une escalade rapide.

En outre, l'utilisation accrue de la cybertechnologie pour des tâches qui étaient autrefois effectuées manuellement ou avec une automatisation de base entraîne le risque que des vulnérabilités de cette technologie soient exploitées à des fins malveillantes. Le recours croissant à des systèmes commerciaux soulève des préoccupations concernant la dépendance à des fournisseurs externes, la perte du contrôle des mises à jour et d'autres inconvénients inhérents aux systèmes exclusifs.

Les points susmentionnés ne sont que quelques exemples des risques que présente l'application de l'IA dans le domaine militaire. À cela s'ajoutent de nombreuses inconnues. Dans un contexte militaire, ces facteurs peuvent, seuls ou associés à d'autres, peuvent compromettre l'issue de la mission et entraîner des risques fondamentaux sur le plan juridique, éthique, humanitaire et militaire.

En outre, des principes clés pour le développement et l'utilisation responsables de l'IA sont énoncés dans la Stratégie norvégienne pour l'intelligence artificielle dans le secteur de la défense (2023). Il s'agit notamment des suivants :

- **Légalité.** Les applications d'IA doivent être développées et utilisées conformément au droit international, y compris le droit international humanitaire et le droit des droits humains. Dans l'étude, le développement, l'acquisition ou l'adoption d'une nouvelle arme, de nouveaux moyens ou d'une nouvelle méthode de guerre reposant sur l'IA, chaque État a l'obligation de déterminer si l'emploi en serait interdit, dans certaines circonstances ou en toutes circonstances, par le droit international des droits humains ou par toute autre règle du droit international applicable à cet État.

- **Responsabilité et obligation de rendre compte**. La responsabilité et l'obligation de rendre compte qui incombent à l'être humain en ce qui concerne l'utilisation de l'IA doivent être garanties. Le pouvoir de décision concernant l'utilisation d'un système d'IA et la responsabilité relative à son emploi réel doivent être définis sans ambiguïté.

- **Explicabilité, intelligibilité, traçabilité**. Les applications d'IA doivent être suffisamment explicables, intelligibles, transparentes et traçables.

- **Formation**. Il est impératif que toute personne travaillant avec l'IA soit dûment formée pour comprendre le fonctionnement de l'application d'IA, notamment pour détecter tout comportement anormal.

- **Fiabilité, sécurité et sûreté**. Les applications d'IA devraient avoir des portées d'utilisation claires et bien définies. Afin de garantir la résilience, la fiabilité et la sécurité de ces applications, il faut faire des essais et des vérifications tout au long de leur cycle de vie, dans le cadre de leurs périmètres d'utilisation respectifs. Les applications d'IA doivent être suffisamment sécurisées et protégées contre les menaces numériques.

- **Contrôle**. Un véritable contrôle humain doit être garanti. Les systèmes d'IA doivent être dotés d'une interface humain-machine adaptée à l'utilisation prévue, qui permette de détecter et d'atténuer les conséquences involontaires, et de prendre les mesures correctrices nécessaires si le système fonctionne de manière inattendue.

Il est nécessaire que la communauté internationale approfondisse le dialogue consacré aux applications de l'IA à des fins militaires et à leurs répercussions sur la paix et la sécurité, notamment en ce qui concerne les dispositions à prendre pour garantir une utilisation responsable de l'IA dans le domaine militaire. Une attention particulière devrait être accordée aux systèmes d'IA utilisés pour appuyer les opérations de combat, notamment pour la perception de la situation et l'aide à la décision, où des résultats et des comportements indésirables dans l'application de

l'IA, et la perte d'un contrôle humain significatif, peuvent avoir des conséquences particulièrement néfastes. Il est également nécessaire d'aborder la question de l'IA dans les guerres hybrides, y compris son emploi dans les cyberopérations, la guerre électronique et les activités d'information.

La Norvège est déterminée à renforcer la coopération internationale en matière de mise en commun de l'information et de renforcement des capacités. En mettant en place une base de connaissances partagée, les États favoriseraient une compréhension commune, combleraient les lacunes, renforceraient la transparence et instaureraient la confiance. À cette fin, la Norvège encouragerait l'élaboration et la publication de stratégies et de documents d'orientation nationaux portant sur les applications de l'IA à des fins militaires. Il convient d'accorder une attention particulière à l'atténuation des risques et aux mesures de confiance.

L'élaboration en temps utile d'une gouvernance internationale adéquate de l'IA, suffisamment souple pour s'adapter aux progrès technologiques rapides, peut aider à prévenir les courses aux armements fondées sur la technologie, tout en veillant à ce que l'innovation contribue à la sécurité mondiale.

Nouvelle-Zélande

[Original : anglais]
[11 avril 2025]

La présente communication nationale de la Nouvelle-Zélande fait suite à la note verbale du Bureau des affaires de désarmement datée du 12 février 2025 et doit être lue en parallèle avec la réponse de la Nouvelle-Zélande à la note verbale du Bureau datée du 1er février 2024[1].

Position de la Nouvelle-Zélande sur l'intelligence artificielle dans le domaine militaire

La Nouvelle-Zélande est consciente que les applications potentielles et existantes de l'intelligence (IA) dans le domaine militaire auront des répercussions considérables et multiformes.

Bien qu'il soit encore difficile de déterminer la nature et l'ampleur de bon nombre de ces répercussions, l'IA est déjà utilisée par certaines organisations militaires dans un large éventail de fonctions militaires, notamment pour le renseignement, la planification, la logistique, la navigation et la communication. L'utilisation de l'IA dans le domaine militaire comporte des risques, mais elle peut offrir des avantages majeurs, notamment en améliorant la rapidité, l'efficacité, la précision et la perception de la situation. À l'instar d'autres forces armées, la Force de défense néo-zélandaise entend tirer parti des possibilités offertes par l'IA pour améliorer ses opérations et maintenir l'interopérabilité avec ses partenaires.

Nous réaffirmons ce qui est dit dans le paragraphe 1 de la résolution 79/239 de l'Assemblée générale, à savoir que « le droit international, notamment la Charte des Nations Unies, le droit international humanitaire et le droit international des droits humains, s'applique aux questions qu'il régit et qui se posent à tous les stades du cycle de vie de l'intelligence artificielle, y compris des systèmes basés sur l'intelligence artificielle, dans le domaine militaire ». Outre les obligations juridiques

[1] Disponible en anglais à l'adresse suivante : https://www.mfat.govt.nz/assets/Peace-Rights-and-Security/Disarmament/New-Zealand-submission-to-the-UN-Secretary-General-on-autonomous-weapon-systems.pdf.

contraignantes, les normes éthiques pertinentes doivent être prises en compte tout au long du cycle de vie de l'IA dans le domaine militaire.

La Nouvelle-Zélande reconnaît que l'IA peut jouer un rôle dans la mise au point et l'emploi de certains systèmes d'armes, notamment en permettant d'accroître leur niveau d'autonomie. Dans sa réponse à la note verbale du Bureau des affaires de désarmement datée du 1er février 2024, elle a présenté en détail sa position sur les systèmes d'armes autonomes.

Il est concevable que l'IA puisse être utilisée pour la mise au point d'armes de destruction massive. Les armes biologiques et chimiques sont clairement interdites par le droit international, et la Nouvelle-Zélande dit que le critère de finalité générale énoncé dans la Convention sur l'interdiction de la mise au point, de la fabrication et du stockage des armes bactériologiques (biologiques) ou à toxines et sur leur destruction et dans la Convention sur l'interdiction de la mise au point, de la fabrication, du stockage et de l'emploi des armes chimiques et sur leur destruction s'appliquerait si l'IA était utilisée pour mettre au point de telles armes, ce qui signifie, entre autres, que l'IA ne doit pas être utilisée à cette fin. En outre, comme l'ont noté les États Parties au Traité sur l'interdiction des armes nucléaires, dont la Nouvelle-Zélande, il est essentiel de maintenir un contrôle humain significatif sur les armes nucléaires et leurs vecteurs, en attendant leur élimination et l'édification d'un monde exempt d'armes nucléaires.

Propositions normatives existantes et nouvelles

Parvenir à une communauté de vues et établir des normes sont des aspects importants de la promotion d'une utilisation militaire responsable de l'IA. En 2024, aux côtés de nombreux autres pays, la Nouvelle-Zélande a adhéré à la Déclaration politique sur l'utilisation responsable de l'intelligence et de l'autonomie artificielles, initiée par les États-Unis. Dans la déclaration, il est dit que l'utilisation militaire de l'IA peut et doit être éthique, responsable et renforcer la sécurité internationale. La Nouvelle-Zélande a également participé aux sommets sur l'intelligence artificielle responsable dans le domaine militaire.

La Nouvelle-Zélande juge utiles les débats multilatéraux, notamment ceux menés dans le cadre de l'Organisation des Nations Unies, consacrés à l'élaboration et à l'adoption des normes relatives à l'IA dans le domaine militaire. Il importe que des acteurs non étatiques, notamment la société civile, les organisations internationales et régionales ainsi que le secteur privé, participent à ces débats à toutes les étapes de ces processus.

Pakistan

[Original : anglais]
[9 avril 2025]

L'essor et l'intégration rapides des technologies d'intelligence artificielle (IA) dans le domaine militaire sont en passe de transformer radicalement les opérations militaires. En effet, l'IA est de plus en plus intégrée dans les systèmes d'armes autonomes, le commandement et le contrôle, les systèmes d'aide à la décision, le renseignement, la surveillance et la reconnaissance, la formation, la logistique et la guerre de l'information ou cyberguerre. Si ces avancées permettent de gagner en efficacité opérationnelle, elles font également peser des risques importants sur la paix et la sécurité internationales.

Difficultés liées à l'intelligence artificielle dans le domaine militaire

Risques stratégiques : interaction avec les armes nucléaires

L'intégration de l'IA aux systèmes d'armes nucléaires fait naître des risques stratégiques, notamment en matière de commandement, de contrôle et de communications nucléaires. Lorsque les capacités d'IA sont intégrées au dispositif de forces nucléaires et aux politiques d'emploi, elles peuvent entraîner des erreurs d'appréciation, des accidents et des conséquences catastrophiques.

Le concept de dissuasion nucléaire repose en grande partie sur la rationalité, la perception et la prise de décision politique humaines. L'intégration de l'IA risque d'éliminer ou de réduire considérablement ces facteurs humains essentiels, ce qui accroît le risque d'escalade automatisée ou accidentelle. Conscients de ces profondes préoccupations, certains États se sont publiquement engagés à conserver un contrôle humain significatif sur les décisions relatives à l'emploi des armes nucléaires, un principe que le Pakistan appuie et qu'il encourage vivement tous les États dotés d'armes nucléaires à adopter.

Dans les régions où des armes nucléaires sont présentes, le recours à des systèmes informatisés d'aide à la décision fondés sur l'IA et à des systèmes d'armes entièrement autonomes dans le domaine conventionnel peut également entraîner des risques d'escalade. Éliminer complètement le contrôle humain en temps de crise pourrait rendre difficile la maîtrise de l'ampleur et de la durée des conflits. Automatiser la riposte dans des contextes volatils où les enjeux sont élevés, en particulier dans des régions où la dynamique nucléaire est tendue, peut aggraver l'enchevêtrement entre les domaines conventionnel et nucléaire et nuire à la stabilité stratégique.

L'utilisation de l'IA pour l'évaluation des données et le renseignement, la surveillance et la reconnaissance peut inspirer une confiance présomptueuse aux États qui envisagent des frappes antiforces préventives et déstabilisatrices ou qui ciblent des capacités de riposte, ce qui fait peser de graves risques sur la stabilité régionale et mondiale.

Risques opérationnels : perte de la maîtrise humaine

L'autonomisation des opérations militaires à l'aide de l'IA risque de diminuer le contrôle humain, ce qui compliquerait la gestion des crises. À mesure que les opérations militaires s'accélèrent pour atteindre la « vitesse de la machine », le temps imparti à la prise de décision humaine se trouve considérablement réduit, ce qui limite les possibilités d'atténuer les crises et de recourir à des solutions diplomatiques.

Les humains pourraient se fier outre mesure aux recommandations que génèrent les systèmes informatisés d'aide à la décision fondés sur l'IA, même si elles sont erronées ou incomplètes, ce qui créerait un biais d'automatisation. Des décisions militaires cruciales pourraient devenir trop dépendantes des résultats produits par les machines, ce qui amènerait les commandants à négliger le contexte, l'intuition ou la prudence propres à l'être humain, et présenterait le risque d'une escalade involontaire des conflits.

L'attrait d'une plus grande efficacité opérationnelle et la quête d'un avantage décisif pourraient inciter à un recours plus fréquent aux capacités fondées sur l'IA, abaissant ainsi le seuil de déclenchement d'un conflit armé. En temps de crise, il serait très déstabilisateur que le seuil de l'emploi de la force soit bas.

Risques techniques

Les applications de l'IA dans le domaine militaire peuvent présenter des failles techniques, comme des biais algorithmiques, l'empoisonnement des données ou une vulnérabilité aux cyberattaques. Des conflits pourraient éclater en raison du mauvais fonctionnement ou de la manipulation des systèmes de détection lointaine ou d'attaques par empoisonnement de données. Les capacités reposant sur l'IA fonctionnent souvent comme des « boîtes noires » et produisent des décisions qui ne sont ni transparentes ni explicables, ce qui rend plus complexes la validation et l'application du principe de responsabilité. De telles vulnérabilités peuvent entraîner des résultats imprévisibles, des défaillances de système et des risques importants pour l'intégrité opérationnelle. Les capacités fondées sur l'IA testées dans un environnement donné avec des jeux de données précis pourraient ne pas fonctionner de manière fiable dans des environnements complètement différents où la dynamique est plus complexe.

Risques normatifs, juridiques et éthiques

Le recours à l'IA dans le domaine militaire soulève des questions d'ordre éthique, normatif et juridique, notamment en ce qui concerne le respect du droit international humanitaire. Le droit international humanitaire repose fondamentalement sur le discernement, le pouvoir discrétionnaire et une prise de décision qui tient compte du contexte, autant de qualités humaines que les systèmes d'IA peuvent difficilement reproduire. Déléguer à des systèmes autonomes des fonctions essentielles, telles que la sélection et l'engagement des cibles, y compris les décisions relatives à la force létale, risque de violer les principes fondamentaux du droit international humanitaire que sont la distinction, la proportionnalité, la précaution dans les attaques et la nécessité militaire. Les systèmes d'IA qui produisent des résultats imprévisibles, peu fiables ou inexplicables compliquent davantage le respect du droit international humanitaire, car ils peuvent entraîner des dommages illégaux ou involontaires.

De plus, l'absence de prise de décision humaine directe ou la dépendance excessive à l'égard des systèmes informatisés d'aide à la décision fondés sur l'IA soulèvent des questions essentielles en matière de responsabilité, ce qui rend extrêmement difficiles l'attribution et la détermination de la responsabilité en cas d'actes illicites ou répréhensibles. En cas de problème, les commandants peuvent rejeter la responsabilité sur l'IA, ce qui compliquerait l'application du principe de responsabilité et l'ouverture d'éventuelles enquêtes sur des crimes de guerre.

Déléguer à des systèmes autonomes des décisions de vie ou de mort soulève d'autres questions éthiques, car cela risque d'amoindrir la compassion, le raisonnement moral et le jugement humain, et donc d'exacerber le risque de violences injustifiées et de victimes civiles.

Prolifération et risques pour la sécurité mondiale

La prolifération des technologies militaires reposant sur l'IA présente des risques importants pour la sécurité internationale. La dissémination de capacités d'IA avancées, en particulier les armes autonomes, risque de provoquer de nouvelles courses aux armements et de déstabiliser la sécurité aux niveaux régional et mondial. Ces craintes sont d'autant plus vives que ces technologies peuvent aisément proliférer et que des acteurs non étatiques peuvent se les procurer.

Réponse internationale proposée : rôle central des mécanismes de l'Organisation des Nations Unies

Les technologies d'IA sont polyvalentes et leurs utilisations pacifiques font partie intégrante de la réalisation des objectifs de développement durable. Dans le même temps, les incidences de l'IA dans le domaine militaire revêtent un caractère transversal et sont susceptibles de peser grandement sur la paix et la sécurité internationales, d'où la nécessité d'agir de manière coordonnée au niveau mondial.

Le Pakistan reconnaît l'intérêt des initiatives de gouvernance de l'IA prises en dehors du système des Nations Unies, mais reste conscient de leurs limites, notamment en ce qui concerne la participation universelle et la légitimité multilatérale officielle. Ces initiatives peuvent certes appuyer l'action menée par l'Organisation des Nations Unies en favorisant le dialogue et la volonté politique, mais les poursuivre de façon isolée risque d'entraîner un éparpillement des efforts. Il convient donc de porter les débats sur les applications militaires de l'IA devant les instances de l'ONU, ce qui permettra de garantir l'inclusivité et la légitimité et de mettre en place un cadre mondial cohérent qui tienne compte des intérêts de tous les États.

Compte tenu des raisons susmentionnées, l'Organisation des Nations Unies doit rester au centre de toute action menée à l'échelle internationale. Il faut que ses mécanismes de désarmement jouent un rôle de premier plan dans les initiatives visant à élaborer un dispositif de gouvernance international relatif à l'IA militaire et à empêcher que le paysage normatif ne se fragmente. Face à l'ampleur et au caractère inédit des répercussions de l'IA dans le domaine militaire, il est nécessaire d'adopter une réponse multilatérale globale et multidimensionnelle. De par sa composition universelle, l'ONU jouit d'une position unique : elle est l'instance idéale où tous les États, qu'ils soient développés ou en développement, peuvent faire entendre leur voix.

Aucune instance, aucun instrument ne saurait suffire à lui seul. Il est nécessaire d'adopter une stratégie structurée qui mette à contribution les différents organes de désarmement de l'ONU, qui travailleront de manière complémentaire, chacun abordant la question sous l'angle qui lui est propre et conformément à son mandat. Le Pakistan propose de tirer parti de toutes les instances compétentes, qu'il s'agisse de l'Assemblée générale et de sa Première Commission, de la Commission du désarmement, de la Conférence du désarmement ou de la Convention sur l'interdiction ou la limitation de l'emploi de certaines armes classiques qui peuvent être considérées comme produisant des effets traumatiques excessifs ou comme frappant sans discrimination. Une telle approche permettrait d'aborder dans tous leurs aspects les dimensions stratégiques, humanitaires, juridiques et techniques, et d'éviter ainsi les lacunes et les doubles emplois. Les travaux de chaque instance devraient éclairer ceux des autres, de façon à créer des synergies pour la réalisation de l'objectif commun, à savoir atténuer les risques liés à l'IA militaire tout en préservant l'utilisation de l'IA à des fins pacifiques.

Conférence du désarmement

La Conférence du désarmement devrait se pencher en priorité sur les risques stratégiques associés à l'IA militaire, en particulier dans le domaine nucléaire, ce qui correspondrait directement aux points 1 et 2 de son ordre du jour (intitulés respectivement « Cessation de la course aux armements nucléaires et désarmement nucléaire » et « Prévention de la guerre nucléaire, y compris toutes les questions qui y sont liées »). En 2023, le Pakistan a proposé d'inscrire un nouveau point à l'ordre du jour de la Conférence sur ce sujet (CD/2334).

Au titre du nouveau point de l'ordre du jour, la Conférence du désarmement devrait créer un organe subsidiaire ou un groupe spécial qui serait expressément chargé d'examiner les risques que l'IA militaire présente pour la stabilité, d'évaluer

la manière dont elle accroît les risques nucléaires et de mener des négociations sur des mesures concrètes à prendre. Ces mesures pourraient consister à :

- prendre l'engagement de maintenir le contrôle humain et de ne pas remplacer le jugement humain dans les décisions relatives à l'emploi des armes nucléaires ;
- interdire l'utilisation de capacités d'IA pour manipuler des données ou cibler les systèmes de commandement, de contrôle et de communication nucléaires ;
- élaborer des mesures de retenue concernant le déploiement et l'utilisation de certaines capacités d'IA susceptibles de déclencher des attaques préventives et d'accroître les risques d'escalade nucléaire.

La Conférence du désarmement est l'instance qui se prête le mieux à de tels débats, puisqu'elle rassemble sur un pied d'égalité tous les États dotés d'importants moyens militaires et que ses décisions sont prises par consensus, ce qui garantit que les intérêts fondamentaux de tous les États en matière de sécurité sont protégés. En se saisissant de la question, elle pourrait donner un nouveau souffle à ses travaux et prouver qu'elle sait faire face aux menaces nouvelles et émergentes.

Commission du désarmement

Forte de sa composition universelle et de son mandat d'organe délibérant, la Commission du désarmement est la mieux placée pour élaborer des directives et des recommandations pratiques sur l'utilisation responsable de l'IA à des fins militaires. Elle a déjà réussi à élaborer des directives similaires (par exemple, sur les mesures de confiance en 1988 et sur les approches régionales du désarmement en 1993).

Dans le cadre de son Groupe de travail II, la Commission du désarmement pourrait élaborer des directives et des recommandations sur des mesures de confiance et de sécurité relatives aux applications militaires de l'IA, aux niveaux mondial et régional. Les éléments clés pourraient consister à réaffirmer les fondements normatifs, à recommander des mesures d'atténuation des risques sur les plans opérationnel et technique, à mettre au point des stratégies visant à atténuer les risques que présente l'IA militaire et à répondre aux préoccupations en matière de prolifération tout en garantissant un accès équitable aux utilisations pacifiques de l'IA.

Première Commission de l'Assemblée générale des Nations Unies

La Première Commission de l'Assemblée générale devrait institutionnaliser l'établissement de rapports d'évaluation périodiques par le Secrétaire général de l'Organisation des Nations Unies et tenir un catalogue des avancées technologiques relatives aux capacités militaires reposant sur l'IA et des risques qui y sont associés, en s'appuyant sur les informations communiquées volontairement par les États Membres. Ces évaluations périodiques donneraient un aperçu fiable de l'évolution des capacités, ce qui permettrait de disposer d'informations actualisées et de faciliter l'adoption de politiques internationales éclairées.

Lors de l'examen des rapports susmentionnés, la Première Commission pourrait tenir des débats consacrés à l'IA et, si nécessaire, créer un groupe de travail à composition non limitée relevant de l'Assemblée générale, en vue de négocier la mise en place d'une plateforme plus institutionnelle, par exemple un registre des Nations Unies sur les applications militaires de l'IA (même si, pour l'instant, il est préférable de tirer parti des instances existantes).

Les rapports mentionnés plus haut pourraient également recenser les domaines dans lesquels un consensus se dégage ou un travail supplémentaire est nécessaire, ce qui permettrait d'orienter les programmes de travail d'instances comme la Conférence

du désarmement, la Commission du désarmement et la Convention sur l'interdiction ou la limitation de l'emploi de certaines armes classiques.

Convention sur l'interdiction ou la limitation de l'emploi de certaines armes classiques

Le Groupe d'experts gouvernementaux créé au titre de la Convention sur certaines armes classiques continue de jouer un rôle primordial dans l'examen des incidences que les systèmes d'armes létaux autonomes ont sur les plans humanitaire, éthique et juridique. Son caractère inclusif (participation de la société civile et du Comité international de la Croix-Rouge en tant qu'observateurs) est un atout.

Le Pakistan salue les travaux menés par le Groupe d'experts gouvernementaux depuis 2017, notamment les 11 principes directeurs établis en 2019. Toutefois, les progrès accomplis dans le cadre de la Convention ont été lents et ont davantage porté sur des principes que sur des réglementations concrètes. Le Pakistan partage l'avis de ceux qui estiment que, dans les débats tenus au titre de la Convention, les aspects sécuritaires des armes dotées de l'IA ont reçu une « attention insuffisante et décroissante », ce qui montre bien qu'il faut que des mesures complémentaires soient prises dans le cadre de la Conférence du désarmement et d'autres instances. Néanmoins, sur le plan humanitaire, le Groupe d'experts gouvernementaux créé au titre de la Convention devrait poursuivre et intensifier ses travaux.

Le Pakistan préconise de conclure des négociations sur un protocole juridiquement contraignant à la Convention qui interdirait les systèmes d'armes létaux autonomes qui fonctionnent sans contrôle humain ou qui ne peuvent être employés dans le respect du droit international humanitaire. Le mandat actuel du Groupe d'experts gouvernementaux autorise les États Membres à élaborer les éléments d'un tel instrument en vue de les présenter à la septième Conférence des Hautes Parties contractantes chargée de l'examen de la Convention, ce qui pourrait permettre d'entamer ensuite des négociations officielles.

Conclusion

Le Pakistan souligne la nécessité de mener une action internationale coordonnée et inclusive pour atténuer les risques importants liés à l'IA militaire. Il préconise une gouvernance permettant de concilier sécurité et développement et de garantir la stabilité, tout en mettant le développement de l'IA au service du bien. En adoptant une stratégie structurée qui fait appel à de multiples instances dans le système des Nations Unies, la communauté internationale peut mettre en place de solides garde-fous normatifs, maintenir la sécurité internationale et préserver un accès équitable et non discriminatoire aux applications pacifiques de l'IA.

Pays-Bas (Royaume des)

[Original : anglais]
[7 avril 2025]

Le Royaume des Pays-Bas se félicite de l'occasion qui lui est donnée de présenter ses vues, comme suite à la résolution 79/239 de l'Assemblée générale adoptée le 24 décembre 2024, sur les possibilités et les difficultés que l'application de l'intelligence artificielle (IA) dans le domaine militaire présente pour la paix et la sécurité internationales.

Les Pays-Bas sont conscients des applications militaires potentielles de l'IA et sont attachés au développement, au déploiement et à l'utilisation responsables de cette technologie dans le domaine militaire. Leur position fondamentale est que l'application de l'IA dans ce domaine doit être conforme au droit international,

notamment à la Charte des Nations Unies, au droit international humanitaire et au droit international des droits humains.

Les 15 et 16 février 2023, les Pays-Bas ont accueilli le premier Sommet sur l'intelligence artificielle responsable dans le domaine militaire. Depuis lors, cette initiative est une instance multipartite qui permet aux représentants de gouvernements, d'institutions du savoir, de cellules de réflexion, de l'industrie et d'organisations de la société civile de débattre des principales possibilités et difficultés liées aux applications militaires de l'IA. Les débats ont lieu chaque année au niveau mondial et tout au long de l'année lors de manifestations régionales sur l'IA responsable dans le domaine militaire, organisées jusqu'à présent par Singapour, le Kenya, la Türkiye, le Chili et les Pays-Bas.

Au sommet de 2023, les Pays-Bas et 57 autres pays ont convenu d'un appel à l'action conjoint sur le développement, le déploiement et l'utilisation responsables de l'IA à des fins militaires. En 2024, les Pays-Bas ont souscrit au plan d'action (Blueprint for Action) adopté lors du Sommet sur l'intelligence artificielle responsable dans le domaine militaire, qu'ils ont coorganisé la même année avec la République de Corée. Ils ont également souscrit à la Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles.

Au sommet de 2023, les Pays-Bas ont institué la Commission mondiale sur l'intelligence artificielle responsable dans le domaine militaire et l'ont chargée de formuler des recommandations à court et à long terme à l'intention des gouvernements et de la communauté multipartite au sens large. Ils attendent la publication du rapport d'orientation stratégique de la Commission mondiale en septembre 2025.

Dans la section ci-après, les Pays-Bas résument leur position et présentent les principales questions qui appellent à un examen plus approfondi.

Possibilités pour la paix et la sécurité internationales

D'un point de vue militaire, les principaux avantages de l'IA sont la rapidité et la portée. L'IA permet un traitement et une analyse bien plus rapides des données. Les systèmes informatisés d'aide à la décision et d'élaboration de scénarios reposant sur l'IA aident également les commandants à créer des plans d'action, ce qui améliore la vision stratégique et la capacité d'intervenir rapidement et efficacement en cas de menaces.

Les Pays-Bas estiment que l'IA a également le potentiel de contribuer à la paix et à la sécurité internationales, car elle peut favoriser une meilleure compréhension de la situation, et permettre d'améliorer la connectivité, de renforcer la protection des civils et d'atténuer les risques lors des opérations de première ligne :

- Les systèmes informatisés d'aide à la décision et d'analyse reposant sur l'IA offrent aux commandants de meilleures informations concernant la situation sur le terrain et l'évolution stratégique à long terme, ce qui contribue à une meilleure compréhension de la dynamique des populations civiles dans les zones de conflit, des difficultés en matière de sécurité climatique, de la violence fondée sur le genre et des modèles de comportement des organisations terroristes. Ces informations peuvent à leur tour être utilisées pour améliorer la gestion des risques et des conflits, renfonçant ainsi la paix et la sécurité internationales.
- Les Pays-Bas estiment que l'IA peut être utile dans le domaine militaire pour améliorer la connectivité entre les forces de défense, ainsi qu'entre ces forces et d'autres acteurs, tels que les acteurs humanitaires, les organismes de surveillance et les administrations locales. Les données peuvent être échangées

entre un grand nombre d'utilisateurs, ce qui crée ainsi des « sources uniques de vérité » avec des capteurs « intelligents » qui fonctionnent dans un environnement sécurisé et en réseau. L'IA peut également être utilisée pour partager des données à des vitesses de plus en plus élevées. L'amélioration de la connectivité grâce à un meilleur partage de données à haut débit contribue à la paix et à la sécurité internationales, car elle renforce la communication, la diffusion de l'information et la coopération internationale, par exemple en ce qui concerne les dispositifs d'alerte rapide et la gestion des crises.

- Les Pays-Bas attachent une grande importance au potentiel de l'IA pour la protection des civils. L'IA est capable de reconnaître les constantes et les écarts dans de grands volumes de données, ce qui peut aider à mieux comprendre l'environnement civil. Cette meilleure compréhension peut réduire le risque d'erreurs d'identification, de dommages collatéraux et de victimes civiles. Plus généralement, l'IA peut aider à mieux reconnaître des menaces pour les civils et les biens de caractère civil, ce qui permet aux forces armées de réagir de manière rapide et appropriée. Elle peut également contribuer à l'amélioration des opérations d'aide humanitaire, telles que la fourniture de nourriture, d'abris et de soins médicaux dans les zones de conflit. Enfin, elle peut faciliter les enquêtes sur les victimes civiles en recueillant et en analysant des données et des preuves, afin de déterminer la cause des dommages et de veiller à ce que les auteurs soient amenés à en répondre.

- L'IA réduit les risques courus par le personnel militaire de première ligne, car les systèmes autonomes reposant sur cette technologie peuvent remplacer les êtres humains dans certaines activités menées dans des environnements difficiles ou dangereux, comme la surveillance sous-marine et l'appui aux opérations de recherche et de sauvetage en conditions météorologiques extrêmes. En réduisant l'exposition du personnel militaire à des environnements à haut risque, l'IA peut également contribuer à diminuer les coûts liés aux soins médicaux et à la réadaptation.

Défis pour la paix et la sécurité internationales

Les Pays-Bas estiment que l'utilisation de l'IA dans le domaine militaire présente plusieurs risques pour la paix et la sécurité internationales.

- Les Pays-Bas craignent que l'IA ne soit utilisée pour amplifier, améliorer et automatiser les cyberattaques et la manipulation de l'information, deux phénomènes qui compromettent la paix et la sécurité internationales. Avec l'essor de l'IA générative, il est plus facile de manipuler l'information et de mener des cyberattaques automatisées. Dans le domaine militaire, de telles activités perturbent les lignes de communication opérationnelles et compliquent la prise de décision. À long terme, la propagation de la désinformation et les cyberattaques automatisées pourraient éroder la confiance dans les lignes de communication militaires. Elles pourraient également miner la confiance entre les États, et donc potentiellement dégrader des relations fragiles, en particulier entre des nations qui sont déjà au bord d'un conflit.

- Les risques associés à l'application de l'IA dans le domaine militaire pourraient conduire à des systèmes susceptibles de violer le droit international. Ces défaillances pourraient résulter d'une mauvaise prise en compte du contexte, des données et du jargon militaire, et entraîner à leur tour une simplification excessive de la prise de décision militaire ou à une méconnaissance des contextes opérationnels précis, par exemple. En outre, les États pourraient violer des obligations juridiques internationales si une application se comporte de manière imprévisible, produit des résultats discriminatoires fondés sur des

caractéristiques non pertinentes ou propose des mesures illégales. En raison de la généralisation croissante de l'IA, le biais d'automatisation, les biais existant dans certains jeux de données ainsi que les décisions humaines fondées sur des systèmes d'IA défaillants risquent de poser de graves problèmes quant à l'attribution des responsabilités, à l'application du principe de responsabilité et à la mise en œuvre de mesures correctives adaptées. Il est important de noter que l'on ne peut pas s'attendre à ce que les applications reposant sur l'IA raisonnent ou fonctionnent de la même manière que les êtres humains.

- Le risque d'escalade induite par l'IA représente une menace pour la paix et la sécurité internationales. Alors que l'IA accélère l'exécution des étapes de la boucle Observation, Orientation, Décision, Action en augmentant la vitesse et la portée des capacités, des perceptions erronées peuvent apparaître en raison des divergences entre les intentions du personnel militaire et les analyses générées par les systèmes reposant sur l'IA. Par conséquent, l'IA pourrait accidentellement contribuer à l'escalade. Les systèmes d'IA étant capables de reconnaître des cibles potentielles à une vitesse et à une distance supérieures à celles des humains, leur utilisation peut également accroître l'intensité et la létalité des conflits.

- Ainsi, la création de systèmes de défense robustes est un défi de plus en plus important. La vitesse avec laquelle les nouvelles applications de l'IA émergent rend difficile la mise en œuvre de stratégies et de tactiques permettant de les contrer et de s'en défendre efficacement dans un contexte militaire. Cette conséquence précise du recours croissant aux systèmes d'IA pourrait favoriser des opérations offensives et, partant, nuire à la paix et à la sécurité internationales.

- La déstabilisation devient une autre préoccupation à mesure que les organisations terroristes, les réseaux de criminalité organisée et d'autres agents non étatiques accèdent à l'IA militaire. Dans ce contexte, les Pays-Bas craignent que l'IA ne soit utilisée par ces acteurs pour faciliter la production d'armes chimiques, biologiques, radiologiques ou nucléaires.

Compte tenu de l'évolution rapide de l'IA, les Pays-Bas reconnaissent qu'il est actuellement impossible de prévoir toutes les possibilités et les difficultés que cette technologie pourrait présenter pour la paix et la sécurité internationales. Certaines sont entièrement nouvelles, tandis que d'autres existent déjà, mais pourraient être exacerbées par l'application de l'IA. Il est essentiel de poursuivre le dialogue international sur cette question afin de veiller à ce que tous les États utilisent l'IA de manière responsable dans le domaine militaire.

Application responsable de l'intelligence artificielle dans le domaine militaire

Afin de garantir une utilisation responsable de l'IA dans le domaine militaire, il faut conserver un discernement et un contrôle humains adaptés au contexte. Les êtres humains doivent rester responsables de l'application de cette technologie et en rendre des comptes. Toutefois, il est important de noter les points ci-dessous.

Un renforcement de la supervision humaine ne garantit pas nécessairement une utilisation plus responsable de l'IA.

Les Pays-Bas estiment qu'il n'existe pas de méthode unique permettant d'assurer un niveau de discernement et de contrôle humains suffisant dans les applications de l'IA. Ce discernement et ce contrôle vont de l'intervention humaine directe à des niveaux plus élevés d'automatisation et d'autonomie, en fonction d'un certain nombre de facteurs. Le degré requis de discernement et de contrôle que les humains devraient exercer sur les applications et les systèmes reposant sur l'IA doit

donc être déterminé au cas par cas. C'est le seul moyen de tenir compte de multiples facteurs, tels que le contexte opérationnel, l'incidence sur la capacité de la technologie à fonctionner de manière autonome dans des environnements hostiles, les paramètres du système et l'interaction humain-machine.

La recherche-développement est essentielle pour un déploiement responsable de l'IA dans le domaine militaire.

Les Pays-Bas croient en l'importance de la recherche-développement. Les États doivent bien évaluer si leurs systèmes d'IA fonctionnent comme prévu et s'ils peuvent être déployés dans un contexte d'utilisation précis. Cela est particulièrement nécessaire lors des combats et dans d'autres environnements où les enjeux sont élevés. Grâce à la recherche-développement en général, ainsi qu'à des procédures éprouvées et fiables d'essai, d'évaluation, de vérification et de validation pour des systèmes d'IA précis, il est possible de détecter et d'éliminer ou d'atténuer les problèmes potentiels avant le déploiement. En outre, il importe que le personnel militaire reçoive une bonne formation et se familiarise bien avec les systèmes d'IA avant leur déploiement, afin de s'assurer qu'il comprend leurs capacités et leurs limites. Cet aspect revêt une importance particulière compte tenu de l'évolution rapide des technologies liées aux systèmes d'IA et du fait que leur utilisation devient de moins en moins coûteuse.

La gouvernance internationale de l'intelligence artificielle militaire devrait être souple, inclusive et réaliste.

En ce qui concerne la gouvernance internationale de l'IA dans le domaine militaire, les Pays-Bas sont conscients qu'il est nécessaire d'adopter une approche souple, équilibrée et réaliste. Premièrement, les cadres de gouvernance doivent être souples afin de pouvoir suivre le rythme rapide de l'évolution des technologies et de la situation sur les champs de bataille. Deuxièmement, les parties doivent s'efforcer de parvenir à une compréhension commune de l'IA dans le domaine militaire, ainsi que des possibilités, des risques et des solutions potentielles qui l'accompagnent. Cela nécessitera un dialogue mondial inclusif et la participation active de tous les groupes de parties prenantes, y compris les États, les institutions du savoir, la société civile et l'industrie. Troisièmement, les États devraient s'attacher à mettre en place des garanties, notamment en matière de traçabilité et d'intelligibilité, afin d'assurer une application responsable de l'IA dans le domaine militaire. Quatrièmement, la gouvernance internationale du déploiement de l'IA militaire doit tenir compte des différents points de vue des États sur la réglementation. Dans les limites des obligations juridiques existantes, cette gouvernance ne devrait pas entraver la capacité des États à innover.

Débat sur les systèmes d'armes autonomes

L'IA ayant un fort potentiel d'utilisation dans les systèmes d'armes autonomes, un parallèle peut clairement être établi entre le débat plus général sur son emploi dans le domaine militaire et le débat sur la réglementation de ces systèmes. Les Pays-Bas considèrent que les discussions menées au niveau international sur ces deux sujets sont complémentaires et utiles d'un côté comme de l'autre.

Pérou

[Original : espagnol]
[11 avril 2025]

Au paragraphe 7 de la résolution 79/239 de l'Assemblée générale des Nations Unies, adoptée le 24 décembre 2024, et pour laquelle le Pérou a voté, le Secrétaire général est prié

> de solliciter les vues des États Membres et des États observateurs sur les possibilités et les difficultés que l'application de l'intelligence artificielle dans le domaine militaire présente pour la paix et la sécurité internationales, en particulier dans des domaines autres que les systèmes d'armes létaux autonomes, et de lui présenter, à sa quatre-vingtième session, un rapport de fond résumant ces vues et répertoriant les propositions normatives existantes et nouvelles, assorti d'une annexe contenant ces vues, dans la perspective de futurs débats entre les États.

Dans ce contexte, le Pérou présente ci-dessous quelques éléments de sa position sur le sujet afin de contribuer à l'élaboration du rapport du Secrétaire général susmentionné.

I. Utilité de l'intelligence artificielle dans le domaine militaire

Le Pérou est conscient de l'évolution rapide et dynamique des technologies émergentes dans le domaine militaire, en particulier pour ce qui concerne les applications potentielles de l'intelligence artificielle (IA). Tout en suivant de près les évolutions dans ce domaine, notamment la manière dont l'IA semble transformer les opérations militaires (de l'utilisation de drones autonomes aux systèmes d'aide à la décision), il apparaît essentiel de promouvoir un dialogue multilatéral soutenu visant à établir des principes garantissant l'utilisation éthique et responsable de ces outils.

Étant donné que l'IA peut être intégrée à la fois aux systèmes d'armes et aux systèmes d'appui aux opérations militaires, le Pérou considère qu'il faut aborder les défis et les préoccupations soulevés par son utilisation d'un point de vue humanitaire, juridique, sécuritaire, technologique et éthique, en tenant compte notamment des risques liés aux biais algorithmiques. Ces inquiétudes sont renforcées par les conséquences que l'utilisation de cette technologie pourrait avoir sur la stabilité et la sécurité internationales.

Cela est d'autant plus grave que les répercussions de l'utilisation de l'IA en matière d'armes nucléaires et d'autres armes de destruction massive risquent d'être de grande portée. Dans ce contexte, il est impératif de renforcer le principe d'un contrôle important par les humains.

II. Vues

Respect du droit international

Le développement, le déploiement et l'utilisation dans le domaine militaire de technologies utilisant l'IA doivent être pleinement conformes au droit international, notamment au droit international des droits humains et au droit international humanitaire, ainsi qu'aux principes fondamentaux inscrits dans la Charte des Nations Unies.

Ainsi, toute élaboration de normes visant à réguler l'IA dans le domaine militaire devrait assurer une utilisation responsable et éthique de celle-ci, ce qui garantit également la non-prolifération des technologies militaires utilisant l'IA et l'accès équitable aux connaissances et aux capacités technologiques.

Il s'agit de veiller à ce que tout recours à l'IA respecte la dignité humaine, protège les civils et garantisse la stabilité et la paix internationales.

Prise en compte des avantages et des risques

L'IA offre de précieuses possibilités de mieux comprendre les situations opérationnelles et donc d'améliorer l'application du droit international humanitaire et la protection des civils et des biens de caractère civil.

Toutefois, son utilisation peut comporter des risques prévisibles et imprévisibles dans le domaine militaire, par exemple en raison de biais algorithmiques, de défauts de conception, d'une mauvaise utilisation ou d'une utilisation malveillante, entre autres. L'IA peut aussi avoir un effet sur des dynamiques régionales et mondiales complexes en influant sur les risques d'escalade, d'erreur d'appréciation, d'abaissement du seuil de déclenchement des conflits ou d'émergence d'une course aux armements.

Développement responsable

L'utilisation de l'IA dans le domaine militaire devrait promouvoir la paix et la protection des civils, et les progrès technologiques devraient compléter, et non remplacer, les capacités humaines.

Comme le préconisent les principes applicables aux systèmes d'armes autonomes, l'utilisation de l'IA dans le domaine militaire doit faire en sorte que la responsabilité et l'obligation de rendre compte ne puissent jamais être transférées aux machines. À cet égard, il faut souligner la nécessité de préserver un contrôle humain marqué dans toute décision liée à l'emploi de la force.

Le Pérou estime que tous les risques et défis liés à cette technologie doivent être traités de manière globale tout au long de son cycle de vie.

Il est possible de mettre en place des contrôles et des garanties pour empêcher l'utilisation abusive de cette technologie dans le domaine militaire sans entraver la recherche, le développement, l'expérimentation et l'innovation liés à l'IA dans d'autres domaines.

Utilisation et transparence

Il faut, en priorité, définir des stratégies, des principes, des normes et des standards, ainsi que des politiques nationales et des cadres juridiques pour garantir l'utilisation responsable de l'IA dans le domaine militaire.

De plus, il importe d'adopter des mesures de confiance et de réduction des risques, et d'instaurer des mécanismes d'échange de bonnes pratiques, dans l'intérêt de la transparence et de la coopération entre les États.

Format des débats

Un dialogue continu doit se poursuivre aux niveaux mondial, régional et interétatique pour ce qui est de l'élaboration de mesures visant à garantir une IA responsable dans le domaine militaire.

Le Pérou est également favorable à une participation inclusive sur cette question, qui tienne compte des positions des États, et en particulier des États en développement, ainsi que de la contribution d'autres parties prenantes, telles que l'industrie, le milieu universitaire, la société civile et les organisations régionales et internationales.

Il importe de ne pas oublier que les États et les régions n'en sont pas tous au même stade d'intégration des capacités d'IA dans le domaine militaire et que leurs environnements de sécurité sont différents.

Dans ce contexte, il semble d'autant plus indispensable d'encourager le renforcement des capacités dans les pays en développement et de consolider la coopération internationale, dans le but de réduire les lacunes existantes et de renforcer la participation de ces pays aux débats sur l'utilisation de cette technologie.

Participation du Pérou aux échanges internationaux

Sommets sur l'intelligence artificielle responsable dans le domaine militaire

- Le Pérou a participé au Sommet en 2023 et 2024, et à l'atelier régional.
- Il a adhéré à la déclaration finale adoptée à l'issue du Sommet en 2024 (« Blueprint for Action »).

Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles

- Le Pérou a participé en tant qu'observateur à la séance plénière inaugurale de cette initiative et a ensuite formalisé son adhésion.
- Sommet pour l'action sur l'intelligence artificielle – Cycle de conférences sur les enjeux de défense (Paris, 2025)
- Le Pérou a envoyé une délégation de haut niveau et souscrit à la Déclaration de Paris sur le maintien du contrôle humain dans les systèmes d'armes dotés d'IA.

République de Corée

[Original : anglais]
[11 avril 2025]

En tant que technologie habilitante, l'intelligence artificielle (IA) a le potentiel de transformer en profondeur de nombreux aspects des affaires militaires, de la prise de décision à la collecte de renseignements, en passant par la logistique, la surveillance et les systèmes de commandement et de contrôle. L'IA se développant rapidement, les États cherchent de plus en plus à l'exploiter dans le domaine militaire.

À mesure que les capacités et les systèmes reposant sur l'IA sont intégrés dans les opérations militaires, ils offrent des possibilités nouvelles, mais posent aussi des difficultés, notamment pour la paix et la sécurité internationales. Cette évolution soulève des questions importantes d'un point de vue humanitaire, juridique, technologique, éthique et de la sécurité.

Aux fins de la présente communication, les vues exposées ci-dessous portent spécifiquement sur des domaines autres que les systèmes d'armes létaux autonomes.

Possibilités offertes par l'IA dans le domaine militaire

Les capacités d'IA et les systèmes dotés de cette technologie, y compris ceux qui sont employés pour le renseignement, la surveillance et la reconnaissance ou comme aide à la décision, permettent d'améliorer la connaissance de la situation, d'accroître la précision, l'exactitude et l'efficacité en traitant des données à grande échelle, en facilitant l'optimisation et en générant des analyses prédictives. Ces capacités et systèmes peuvent contribuer à maintenir et à promouvoir la paix et la sécurité internationales.

1. Renforcer l'application du droit international humanitaire et contribuer à protéger les civils et les biens de caractère civil dans les conflits armés

Les systèmes de renseignement, de surveillance et de reconnaissance et les systèmes informatisés d'aide à la décision fondés sur l'IA peuvent renforcer l'application des principes fondamentaux du droit international humanitaire, à savoir les principes de distinction, de proportionnalité et de précaution dans les attaques, car ils permettent de mieux évaluer la situation sur le champ de bataille et de mieux la comprendre. L'IA peut aider à distinguer les combattants des non-combattants et à évaluer les dommages collatéraux potentiels, en s'appuyant sur des informations opportunes et circonstanciées. En permettant de mieux connaître la situation sur le champ de bataille, et notamment de savoir si des civils s'y trouvent, l'IA aide à juger s'il est nécessaire et opportun de prendre des précautions pour protéger les civils et les infrastructures civiles.

2. Appuyer les opérations de maintien de la paix

L'IA peut aider à surveiller l'application des accords de cessez-le-feu et des accords de paix. Elle peut aussi appuyer les dispositifs d'alerte rapide visant à déceler d'éventuelles violations, ce qui renforce l'efficacité et la sûreté des missions. La République de Corée a lancé un projet pilote de « camp intelligent » au sein de l'unité Hanbit de la Mission des Nations Unies au Soudan du Sud (MINUSS) afin de tirer parti de l'IA et d'autres technologies émergentes pour améliorer la sécurité, l'efficacité et les capacités opérationnelles des camps de maintien de la paix des Nations Unies.

3. Renforcer les capacités de vérification et de contrôle des régimes de maîtrise des armements et de conformité

L'IA peut renforcer les capacités des mécanismes de vérification internationaux visant à contrôler le respect des accords de maîtrise des armements et de non-prolifération. L'Agence internationale de l'énergie atomique (AIEA) pourrait tirer parti de l'IA pour accroître l'efficacité des processus de garanties, en particulier pour ceux qui consistent notamment à classer des données, à trouver des constantes et à repérer des valeurs aberrantes. Les systèmes fondés sur l'IA peuvent également aider à déceler les premiers indices de l'emploi d'armes chimiques ou biologiques et à mettre au jour des tactiques de plus en plus sophistiquées visant à se soustraire aux sanctions, ce qui consolide les régimes internationaux de non-prolifération.

Outre les possibilités décrites ci-dessus, l'IA peut contribuer à atténuer les risques stratégiques, comme les erreurs d'appréciation, les malentendus et l'escalade involontaire, car elle permet de mieux analyser le comportement des acteurs et d'être mieux à même de détecter les menaces et d'y répondre de manière proactive. De plus, les capacités d'IA peuvent aider à se doter de moyens de renforcer la cyberdéfense, de protéger les infrastructures nationales essentielles et de combattre le terrorisme, entre autres.

Difficultés posées par l'intelligence artificielle dans le domaine militaire

Si elle n'est pas conçue, déployée et utilisée de manière responsable, l'IA militaire risque de créer de nouvelles difficultés ou d'exacerber celles qui existent déjà.

Les difficultés peuvent tenir aux caractéristiques techniques et opérationnelles de l'IA. Le fonctionnement de l'IA, assimilable à celui d'une « boîte noire », complique la compréhension de la manière dont des résultats précis sont générés et les raisons qui les sous-tendent, ce qui limite leur explicabilité et leur traçabilité. Des défauts de conception et des biais involontaires dans les données, les algorithmes ou

l'architecture informatique peuvent entraîner des dysfonctionnements ou des résultats qui s'écartent des objectifs visés. Une confiance excessive dans les systèmes d'IA, qui se traduit notamment par le biais d'automatisation, ou une formation insuffisante peut poser problème en raison de l'absence d'une intervention et d'un jugement humains appropriés. Ces facteurs pourraient accroître le risque d'erreur d'appréciation, d'interprétation erronée ou d'escalade involontaire en situation de conflit, ce qui mettrait en péril la paix et la sécurité internationales.

La nature à double usage des technologies d'IA pourrait accroître le risque d'utilisation abusive ou de détournement par des acteurs irresponsables animés d'intentions malveillantes. Par exemple, dans le domaine cyber, les campagnes de désinformation et les cyberattaques reposant sur l'IA, telles que l'empoisonnement des données et l'usurpation d'adresse, pourraient s'accélérer. En outre, des acteurs irresponsables pourraient exploiter les technologies d'IA pour faciliter la mise au point de nouvelles armes chimiques ou biologiques, ce qui soulève des préoccupations en matière de prolifération et augmente les risques pour la paix et la sécurité internationales.

Application de l'intelligence artificielle responsable dans le domaine militaire

Afin de tirer parti des avantages et des possibilités offerts par l'IA tout en prenant en compte les risques et les difficultés qui y sont associés, les capacités d'IA et les systèmes qui en dépendent dans le domaine militaire doivent être développés, déployés et utilisés de manière responsable tout au long de leur cycle de vie.

La République de Corée s'engage à assurer et à promouvoir une application responsable de l'IA dans le domaine militaire. Cet engagement repose sur les principes et mesures clés suivants :

- L'IA doit être éthique et centrée sur l'être humain.
- Les capacités d'IA dans le domaine militaire doivent être appliquées conformément au droit international applicable, y compris le droit international humanitaire et le droit international des droits humains.
- La responsabilité et l'obligation de rendre compte de l'utilisation de l'IA dans le secteur militaire et de ses conséquences incombent toujours aux êtres humains, et ne peuvent en aucun cas être transférées à des machines.
- La fiabilité des applications d'IA doit être assurée par la mise en place de garde-fous appropriés afin de réduire les risques de dysfonctionnements ou de conséquences imprévues, notamment ceux découlant des biais liés aux données, aux algorithmes et autres.
- Une intervention humaine adéquate doit être maintenue dans le développement, le déploiement et l'utilisation de l'IA dans le domaine militaire, y compris par l'adoption de mesures adéquates visant à garantir que le jugement et le contrôle relatifs à l'emploi de la force demeurent exercés par des êtres humains.
- Le personnel concerné devrait être en mesure de comprendre, d'expliquer, de retracer les résultats produits par les capacités d'IA et de faire confiance à ces résultats dans le domaine militaire, ainsi que les systèmes qui reposent sur cette technologie. Il convient de poursuivre les efforts visant à accroître l'explicabilité et la traçabilité des résultats générés par l'IA.

La République de Corée appuie les débats et dialogues destinés à promouvoir davantage de mesures pour garantir une IA responsable dans le domaine militaire, notamment par le biais de cadres normatifs internationaux, de protocoles d'essai et d'évaluation rigoureux, de processus complets de vérification, de validation et d'accréditation, de mécanismes nationaux de suivi solides, de processus de contrôle

continu, de programmes de formation et d'exercices complets, d'une cybersécurité renforcée, de dispositifs clairs d'application du principe de responsabilité.

Il est essentiel de mettre en place des mesures robustes de contrôle et de sécurité pour empêcher les acteurs irresponsables d'acquérir et d'utiliser à mauvais escient des capacités d'IA potentiellement nuisibles dans le domaine militaire, y compris les systèmes qui reposent sur cette technologie.

La République de Corée encourage l'élaboration de mesures de confiance efficaces et de mesures appropriées de réduction des risques, ainsi que l'échange d'informations et les consultations sur les bonnes pratiques et les enseignements tirés entre les États.

La République de Corée insiste sur la nécessité d'empêcher que les capacités d'IA ne soient utilisées pour contribuer à la prolifération des armes de destruction massive par des acteurs étatiques et non étatiques et souligne que ces capacités ne devraient pas entraver les efforts de maîtrise des armements, de désarmement et de non-prolifération. Il est crucial de maintenir le contrôle humain et l'intervention humaine à toutes les actions déterminantes visant à informer et à exécuter les décisions souveraines concernant l'emploi des armes nucléaires, sans préjudice de l'objectif ultime d'un monde exempt d'armes nucléaires.

Il importe de développer, de déployer et d'utiliser les capacités d'IA et les systèmes reposant sur l'IA dans le domaine militaire de manière à maintenir et à ne pas entraver la paix et la sécurité internationales.

Gouvernance future de l'intelligence artificielle dans le domaine militaire

Envisager la gouvernance future de l'IA militaire implique de promouvoir une compréhension commune de cette technologie, de ses atouts et de ses limites, ainsi qu'une compréhension partagée de ses applications potentielles dans ce secteur et de leurs répercussions pour la paix et la sécurité internationales.

Il importe également de renforcer les capacités, notamment dans les pays en développement, afin de promouvoir leur pleine participation aux débats sur la gouvernance et de faciliter une approche responsable ainsi qu'une compréhension commune du développement, du déploiement et de l'utilisation de l'IA dans le domaine militaire. L'échange de connaissances, de bonnes pratiques et d'enseignements tirés de l'expérience peut également contribuer à une compréhension commune.

Compte tenu des progrès rapides de l'IA, les mécanismes de gouvernance doivent être suffisamment souples pour s'adapter à son évolution. La République de Corée est également favorable à une approche équilibrée qui tienne compte à la fois des possibilités et des risques. Des discours de gouvernance trop axés sur les risques ou trop restrictifs pourraient entraver l'innovation et occulter le potentiel de l'IA à contribuer à la paix et à la sécurité internationales. La gouvernance future ne devrait pas constituer un obstacle à l'innovation, mais plutôt l'appuyer et faciliter une application responsable de l'IA dans le domaine militaire.

Étant donné que la communauté internationale n'en est qu'aux premiers stades de la compréhension des répercussions de l'IA militaire pour la paix et la sécurité internationales et compte tenu de l'état actuel du développement technologique et de l'évolution des politiques, il serait prématuré de définir de manière restrictive la trajectoire de la gouvernance de l'IA ou d'établir des instruments ou des normes juridiquement contraignants sans une compréhension commune et partagée de ce qu'est l'IA responsable dans le domaine militaire. La République de Corée estime que les débats sur la gouvernance doivent être réalistes et se dérouler progressivement, dans le cadre d'un dialogue continu.

Consciente que l'innovation en matière d'IA est portée par le secteur privé, la République de Corée estime que les futurs efforts de gouvernance doivent adopter une approche ouverte et inclusive mobilisant de multiples parties prenantes, notamment l'industrie, le monde universitaire, la société civile et les organisations régionales et internationales.

La République de Corée salue les efforts entrepris aux niveaux national, régional et mondial pour saisir les possibilités et relever les défis liés à l'IA dans le domaine militaire, notamment l'élaboration de stratégies, de lois, de principes, de normes, de politiques et de mesures à l'échelle nationale. Elle reconnaît qu'il importe de promouvoir le dialogue à tous les niveaux.

Afin de veiller à l'application responsable de l'IA dans le domaine militaire, la République de Corée a créé en 2022 la Division des politiques des données, puis, en 2025, l'équipe des politiques d'intelligence artificielle en matière de défense au sein du Ministère de la défense nationale. En 2024, le Ministère a mis en place le Comité des données et de l'intelligence artificielle en matière de défense en tant qu'organe de délibération et de décision au plus haut niveau.

Afin de promouvoir le dialogue, en collaboration avec les Pays-Bas, Singapour, le Kenya et le Royaume-Uni, la République de Corée a organisé le deuxième Sommet sur l'intelligence artificielle responsable dans le domaine militaire en septembre 2024 à Séoul. Les Sommets sur l'intelligence artificielle responsable dans le domaine militaire et une série de consultations régionales tenues sur la question en 2024 ont offert un cadre propice à l'échange de compétences et à la promotion d'un dialogue inclusif et d'une compréhension mutuelle. À l'avenir, le troisième Sommet sur l'intelligence artificielle responsable dans le domaine militaire, qui se tiendra en Espagne en septembre 2025, ainsi que les prochaines consultations régionales sur l'IA responsable dans le domaine militaire prévues en 2025, continuera de guider l'action que mène la communauté internationale pour promouvoir une application responsable de l'IA dans le domaine militaire.

La République de Corée estime que les débats sur l'application responsable de l'IA dans le domaine militaire qui ont lieu dans le cadre de l'Organisation des Nations Unies, y compris la Première Commission de l'Assemblée générale et la Commission du désarmement, devraient s'inscrire en complémentarité avec d'autres initiatives pertinentes menées en dehors de l'ONU, y compris le processus des Sommets sur l'intelligence artificielle responsable dans le domaine militaire, la Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles, et le Groupe d'experts gouvernementaux sur les technologies émergentes dans le domaine des systèmes d'armes létaux autonomes. Elle considère que ces initiatives se renforcent mutuellement et sont complémentaires.

La gouvernance des données est également cruciale. Étant donné que les données jouent un rôle central dans la formation, le déploiement et l'évaluation des systèmes reposant sur l'IA, il est impératif que les parties prenantes concernées s'engagent dans des discussions approfondies sur des mécanismes de gouvernance des données adéquats, y compris des politiques et procédures claires pour la collecte, le stockage, le traitement, l'échange et la suppression des données, ainsi que la protection des données.

Royaume-Uni de Grande-Bretagne et d'Irlande du Nord

[Original : anglais]
[11 avril 2025]

L'intelligence artificielle (IA) est un ensemble de technologies à usage général, dont chacune peut permettre à des machines d'effectuer des tâches qui nécessiteraient traditionnellement une intelligence humaine ou biologique, en particulier lorsque les machines apprennent à partir de données comment effectuer ces tâches. Ces technologies arrivent à maturité et sont adoptées à un rythme exceptionnel. Chacune d'entre elles a ses propres systèmes, méthodes et applications, et présente donc des trajectoires de développement distinctes ainsi que des incidences différentes. Il est certain qu'elles peuvent entraîner des changements profonds dans tous les aspects de la société, de l'économie et de la politique, y compris la défense et la sécurité.

Le Royaume-Uni se félicite de l'occasion offerte par la résolution 79/239 de l'Assemblée générale d'examiner les répercussions de l'IA dans le domaine militaire au-delà de celles liées aux systèmes d'armes létaux autonomes, qui ont fait l'objet de débats approfondis et précieux, notamment dans le Groupe d'experts gouvernementaux créé dans le cadre de la Convention sur l'interdiction ou la limitation de l'emploi de certaines armes classiques qui peuvent être considérées comme produisant des effets traumatiques excessifs ou comme frappant sans discrimination. Il importe de procéder à une évaluation rigoureuse des incidences stratégiques plus vastes de l'IA militaire, en s'appuyant sur les réflexions, idées et meilleures pratiques relatives à cette question, telles qu'elles sont abordées dans les instances internationales informelles et formelles, ce qui permettra d'engager une discussion globale sur la manière de tirer le meilleur parti des possibilités qu'offre l'IA dans le domaine militaire, tout en atténuant efficacement les risques associés.

Possibilités offertes par l'IA dans le domaine militaire

L'intégration de l'IA dans le domaine militaire transformera sans doute la défense, la dynamique de la sécurité mondiale et la nature de la guerre. Le recours à des technologies de pointe reposant sur l'IA, capables de classer et d'affiner plus rapidement et de façon plus exhaustive de vastes quantités de données issues de sources variées, permettra d'accroître l'efficacité, d'améliorer la prise de décision, d'accélérer le rythme de la planification opérationnelle et d'en accroître la rigueur. Intégrée aux systèmes de renseignement, de surveillance et de reconnaissance, l'IA peut offrir une vision plus précise du contexte opérationnel et aider les planificateurs à réduire l'impact sur les civils, assurant ainsi une meilleure protection des personnes et des infrastructures civiles. L'automatisation de la logistique et de la gestion des engins non explosés réduira la nécessité d'avoir du personnel militaire sur le terrain. Ainsi, l'IA militaire pourrait contribuer à renforcer la sécurité nationale et internationale, à réduire les risques pour la vie humaine et à diminuer le nombre de victimes.

Les recherches menées par le Ministère britannique de la défense sur l'IA et le maintien de la paix ont permis de cerner les moyens par lesquels les opérations de paix pourraient tirer parti des capacités et systèmes optimisés par l'IA, par exemple :

- une capacité analytique qui améliorera la perception de la situation, la prise de décision opérationnelle, la planification de scénarios et la capacité d'analyse de sentiments ;
- des systèmes autonomes, tels que les véhicules aériens sans pilote, pourraient améliorer la couverture de zones étendues ou de régions à haut risque, où la présence permanente du personnel de maintien de la paix serait dangereuse ;

- l'optimisation de la logistique pourrait faciliter la prestation de soins de santé et la fourniture d'aide aux populations locales, ce qui viendrait appuyer les objectifs des missions de maintien de la paix et renforcer la confiance que ces populations leur témoignent ;

Les capacités susmentionnées peuvent servir à mieux surveiller et suivre l'application des accords de paix et de maîtrise des armements, afin de détecter plus facilement les violations ou de confirmer, en temps voulu et de manière crédible, le respect de ces accords. Grâce aux outils d'IA, on pourrait mieux détecter, identifier, attribuer et vérifier les opérations hostiles de diverses sortes qui se trouvent en deçà du seuil de conflit armé, ce qui réduirait l'efficacité de ces activités et pourrait même en dissuader les auteurs. Ils peuvent également aider à reconnaître et à surveiller en temps réel les discours haineux, la propagande ou l'évolution de l'opinion publique en ligne, qui sont susceptibles de faire monter les tensions, de saper les pourparlers de paix ou de compromettre le cessez-le-feu.

Défis et risques

L'utilisation de l'IA dans le contexte militaire peut exacerber les risques existants et créer de nouvelles menaces, aussi bien au-dessus qu'en deçà du seuil de conflit armé. Dans la course à l'adoption des capacités d'IA dans le but d'obtenir un avantage stratégique, les pays pourraient recourir à l'IA de manière inacceptable sur le plan juridique, éthique ou de la sécurité. Il faudra de nouveaux protocoles et mécanismes de désescalade pour gérer les risques d'escalade ou d'accident liés à l'IA et résultant des dysfonctionnements ou de la fragilité, de la vulnérabilité, de l'immaturité ou des failles de sécurité des systèmes fondés sur l'IA. Des acteurs hostiles peuvent chercher à attaquer les systèmes nationaux reposant sur l'IA et à miner la confiance dans leurs performances, leur sécurité et leur fiabilité (par exemple, en « empoisonnant » les sources de données, en corrompant des composants matériels dans les chaînes d'approvisionnement et en perturbant les communications et les commandes), ce qui pourrait déstabiliser les systèmes et fausser la prise de décision militaire en temps de crise ou dans d'autres contextes opérationnels.

Grâce au rythme opérationnel qu'elles permettent, en période de conflit, ces technologies sont susceptibles de réduire drastiquement les délais de décision, de mettre à rude épreuve les capacités de compréhension humaine et d'exiger des réponses à la vitesse de la machine. Étant donné que de nombreuses capacités d'IA fonctionnent comme une « boîte noire », les êtres humains sont souvent incapables de discerner comment ou pourquoi un résultat particulier a été produit. Les opérations fondées sur l'IA peuvent donner lieu à des comportements imprévisibles et opaques et rendre difficiles les déductions et les jugements précis sur les intentions d'un adversaire, voire être mal interprétées ou provoquer des conséquences involontaires. Les opérateurs pourraient faire une confiance excessive aux résultats algorithmiques sans comprendre pleinement les hypothèses, les contraintes et les défauts sous-jacents des systèmes reposant sur l'IA. En l'absence de garanties, de normes et de protocoles appropriés, ces systèmes reposant sur l'IA pourraient exacerber le risque d'incompréhension, d'erreur d'appréciation et d'escalade involontaire.

La disponibilité généralisée de capacités et d'outils d'IA avancés et d'autres technologies à double usage est susceptible d'accroître les risques de prolifération et la mise au point de nouvelles armes par des acteurs étatiques ou non étatiques. L'IA pourrait également être utilisée pour intensifier les tentatives de désinformation visant à susciter l'hostilité envers certains pays, ce qui pourrait entraîner des conflits et exacerber les tensions.

Engagement du Royaume-Uni en faveur d'une intelligence artificielle sûre et responsable dans le domaine militaire

Le Royaume-Uni reconnaît que l'IA soulève de profondes préoccupations concernant l'équité, les biais, la fiabilité et la nature de la responsabilité humaine et de l'obligation de rendre compte, en particulier dans un contexte militaire. Les États ont une longue histoire d'intégration de nouvelles technologies et continueront de s'appuyer sur des régimes juridiques, de sécurité et réglementaires établis de longue date. Cependant, nous devons reconnaître les défis particuliers découlant de la nature de l'IA et l'importance de démontrer que nous sommes responsables et dignes de confiance.

Par sa stratégie de défense en matière d'IA et des principes éthiques qui y sont associés, le Royaume-Uni s'engage en faveur d'une IA sûre et responsable. Ces principes en matière d'IA, énoncés dans la politique britannique « Ambitious, Safe and Responsible » (ambitieuse, sûre et responsable), servent de fondement à un cadre structuré autour de cinq considérations : la primauté de l'être humain, la responsabilité, la compréhension, la réduction des biais et des préjudices et la fiabilité. Parue en novembre 2024, la publication interarmées « Dependable Artificial Intelligence (AI) in Defence » (pour une intelligence artificielle fiable dans le domaine de la défense) présente des orientations claires destinées aux équipes du Ministère de la défense et au-delà sur la manière d'appliquer ces principes éthiques pour fournir des services et capacités fondés sur l'IA, robustes, fiables et efficaces.

Par ses principes éthiques en matière d'IA, le Royaume-Uni entend renforcer la confiance dans les technologies et applications de l'IA, exploiter pleinement le potentiel de la collaboration humain-machine, tout en réduisant les risques liés à leur usage, à leur mésusage ou à l'absence d'usage, et en prévenant toute conséquence involontaire. Cette approche permet au Royaume-Uni d'exploiter l'innovation et la créativité que l'on trouve dans les secteurs de la défense et de l'industrie d'une manière qui facilitera l'adoption ambitieuse de solutions fondées sur l'IA.

Le Royaume-Uni indique clairement que toute utilisation de l'IA pour améliorer ses processus de défense, ses systèmes ou ses capacités militaires est régie par le droit national et international. Les forces armées du Royaume-Uni s'efforcent toujours de respecter leurs obligations juridiques dans toutes leurs activités, du droit de l'emploi à la protection de la vie privée et aux marchés publics, en passant par le droit des conflits armés, également connu sous le nom de droit humanitaire international. Elles disposent de pratiques et de processus rigoureux pour s'assurer que leurs activités et leur personnel respectent la loi. Ces pratiques et processus s'appliquent et continueront de s'appliquer aux capacités reposant sur l'IA. Le déploiement de ces capacités dans les conflits armés doit être pleinement conforme au droit international humanitaire, et satisfaire aux quatre principes fondamentaux de distinction, de nécessité, d'humanité et de proportionnalité. Il est clair que l'utilisation de tout système ou arme qui ne respecte pas ces principes fondamentaux constituerait une violation du droit international.

Il est également essentiel que la responsabilité et l'obligation de rendre compte qui incombent aux êtres humains s'exercent par une intervention humaine adaptée au contexte. Cette intervention humaine est nécessaire pour respecter nos politiques, nos principes éthiques et nos obligations découlant du droit humanitaire international. La nature de l'intervention humaine variera en fonction de la nature de la capacité, de l'environnement opérationnel et du contexte d'utilisation. Le Royaume-Uni veillera au maintien d'un contrôle politique exercé en permanence par des décideurs humains sur ses armes nucléaires.

Contribution du Royaume-Uni aux initiatives internationales

Pour la stabilité mondiale, il faut que l'IA militaire soit développée de manière ambitieuse, mais responsable. La communauté internationale comprend de mieux en mieux les risques, les garanties et les normes liés à l'utilisation de l'IA dans le contexte militaire. Étant donné que les risques sont par nature internationaux, ils nécessitent une réponse mondiale.

Le Royaume-Uni est à l'avant-garde des initiatives internationales visant à promouvoir un développement et une utilisation sûrs et responsables de l'IA. Le Royaume-Uni se félicite d'avoir accueilli le premier Sommet sur la sécurité de l'IA, qui a abouti à l'adoption de la Déclaration de Bletchley sur la sécurité de l'IA, et d'avoir aidé à commander le Rapport international sur la sécurité de l'IA, première synthèse mondiale des textes existants sur les risques et capacités des systèmes d'IA avancés. Publié en février 2025, ce rapport permet de mieux comprendre les éléments essentiels pour éclairer les débats internationaux, notamment ceux portant sur l'IA au service de la paix et de la sécurité. Il appuie les efforts déployés dans le cadre du Pacte numérique mondial afin de réduire la fracture numérique et d'améliorer la gouvernance internationale en matière d'IA au profit de l'humanité.

Le Royaume-Uni appuie activement les initiatives internationales visant à stimuler l'action dans le domaine militaire. Il a soutenu les travaux menés par des organisations comme RAND Europe, l'Université de Californie à Berkeley et la Commission mondiale sur l'intelligence artificielle responsable dans le domaine militaire, afin de réunir divers experts de renom pour examiner ces questions, étudier les réflexions les plus récentes et formuler des recommandations concrètes à l'attention des décideurs politiques. Le Royaume-Uni poursuit sa participation active aux dialogues internationaux sur les questions de défense et de sécurité liées à l'IA et continue de partager son expérience en matière d'élaboration et d'application d'approches sûres et responsables pour l'adoption de l'IA dans le domaine militaire. Il se félicite des progrès accomplis grâce à des initiatives comme les Sommets sur l'intelligence artificielle responsable dans le domaine militaire, qu'il a coorganisés en 2024, et la Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles, élaborée sous l'égide des États-Unis. Ces efforts visent à approfondir la compréhension des possibilités et des risques stratégiques, et à y répondre par des mesures appropriées qui favorisent une utilisation sûre et responsable de l'IA. L'éthique et les garanties en matière d'IA sont des domaines dynamiques qui exigent un engagement, une collaboration et une itération continus.

Perspectives

Le Royaume-Uni est impatient de tirer parti des progrès accomplis à ce jour dans le cadre des processus existants, notamment lors des débats à l'ONU sur le rapport du Secrétaire général, axés sur des actions concrètes. Compte tenu de la nature de l'IA militaire, il sera essentiel d'adopter une approche inclusive et multipartite, fondée sur les compétences techniques, militaires et juridiques des États, de l'industrie, du monde universitaire et de la société civile.

Nous disposons d'une abondance d'informations, mais notre compréhension collective des applications et des répercussions de l'IA dans le domaine militaire reste faible, et il subsiste des lacunes et des malentendus importants quant à la nature et aux capacités de l'IA. Des travaux supplémentaires sont nécessaires pour renforcer les capacités des États, approfondir notre compréhension collective des répercussions, risques et défis stratégiques de l'IA militaire et établir une terminologie universellement acceptée pour permettre des échanges constructifs. Les débats devraient se concentrer sur des mesures et des pratiques concrètes, efficaces et

appropriées qui pourraient contribuer à atténuer les risques, notamment des garanties et des normes de comportement, de nouveaux canaux de communication et des mécanismes de transparence visant à réduire le risque de mauvaise interprétation, des doctrines actualisées, des mesures de confiance et des accords de maîtrise des armements qui tiennent compte de l'incidence de l'IA militaire.

Serbie

[Original : anglais]
[4 avril 2025]

Le développement et l'application de l'intelligence artificielle (IA) constituent un facteur important de changement dans la manière dont les opérations militaires sont menées dans le monde d'aujourd'hui. L'IA offre de nouvelles possibilités, qui s'accompagnent néanmoins de défis nouveaux pour la stabilité internationale, la paix et la sécurité dans le domaine militaire. Il est donc nécessaire d'entreprendre la création d'un cadre international approprié pour réglementer son application.

1. Possibilités et avantages de l'utilisation de l'intelligence artificielle dans le domaine militaire

Dans un contexte militaire, recourir à l'IA à des fins non létales peut contribuer à améliorer de nombreux aspects des opérations militaires :

a) Améliorer le niveau de connaissance des opérations ;

b) Améliorer la qualité et la rapidité de la prise de décision ;

c) Améliorer la qualité des données de renseignement et de reconnaissance par un traitement rapide des données et permettre une détection rapide des menaces ;

d) Appuyer la protection des civils et des non-combattants dans les conflits militaires ;

e) Appuyer les opérations et missions de paix en facilitant le suivi de l'application des accords de cessez-le-feu et en prédisant la dynamique des conflits ;

f) Améliorer les processus et les procédures de maintenance prédictive et d'optimisation logistique en favorisant la réduction des coûts et l'économie des ressources.

2. Principaux défis et risques liés à l'utilisation de l'intelligence artificielle dans le domaine militaire

Le développement et l'intégration de l'IA dans les systèmes de combat et les systèmes non destinés au combat posent un défi majeur pour la paix et la stabilité internationales, et pour le droit international humanitaire, principalement dans les domaines suivants :

a) Les risques techniques et les défaillances de fonctionnement dus à des erreurs d'application dans un environnement dynamique, qui peuvent menacer des vies humaines, causer des dommages matériels et influer sur la mise en œuvre du droit international humanitaire ;

b) Les risques juridiques et éthiques liés au respect du droit international, notamment en ce qui concerne l'application de ses principes, tels que la distinction, la proportionnalité et les mesures de précaution en matière de ciblage ;

c) L'absence de règles explicites permettant d'établir la responsabilité des actes et activités faits par l'IA ;

d) L'imperfection des algorithmes risque de conduire à des biais, de la discrimination et des erreurs dans le processus de prise de décision, car l'utilisation de jeux de données non représentatifs peut entraîner une identification erronée de civils ou constituer une menace pour certains groupes ethniques ou nationaux ;

e) L'utilisation des algorithmes d'IA peut donner aux personnes qui prennent part à la conduite des opérations l'illusion d'une moindre responsabilité ;

f) Les risques stratégiques liés aux décisions prises par l'IA et fondées sur des prémisses erronées ;

g) La convergence et l'intégration non sélectives avec les nouvelles technologies, en particulier dans les domaines de l'information et des cyberopérations ou de l'utilisation de moyens nucléaires, chimiques et biologiques ;

h) Le manque de personnel qualifié capable de développer, de déployer et d'utiliser de manière responsable les systèmes d'IA en situation de conflit ;

i) L'utilisation malveillante de l'IA dans les opérations d'information, par la création et la diffusion de la désinformation, ce qui peut inciter à des conflits et exacerber les tensions.

3. Créer un cadre juridique et éthique

Compte tenu des risques et difficultés évalués, il est nécessaire de mettre en place, au sein de la communauté internationale, des cadres juridiques et éthiques contraignants visant à faire ce qui suit :

a) Promouvoir l'ouverture d'un dialogue au sein de l'ONU et y contribuer, afin d'accroître le respect des normes du droit international humanitaire, y compris l'établissement de normes, règles et principes juridiques internationaux qui garantiraient que le développement et l'application de systèmes reposant sur l'IA sont conformes aux principes du droit international humanitaire (distinction, proportionnalité et mesures de précaution visant à protéger les personnes civiles) ;

b) Engager un processus d'évaluation de la légalité de l'emploi des systèmes et des armes au regard des applications approuvées de l'IA ;

c) Assurer la protection de la vie et de la liberté des individus pendant les conflits armés et de leur vie privée en temps de paix, en particulier dans le contexte de la surveillance ;

d) Consolider les mécanismes des Nations Unies en introduisant l'examen obligatoire des risques liés à l'application de l'IA à des fins militaires, en renforçant la Conférence du désarmement, en harmonisant les travaux de la Commission du désarmement, en créant de nouveaux organes spécialisés des Nations Unies et en élargissant les initiatives existantes des Nations Unies pour l'utilisation responsable de l'IA ;

e) Entamer un dialogue au sein de l'ONU pour définir l'utilisation responsable de l'IA dans le domaine militaire et établir des protocoles de sécurité pour son application (essais, évaluation, validation et vérification) ;

f) Élaborer des mesures visant à harmoniser la contribution du secteur privé avec les principes du droit humanitaire international lors de la conception, du déploiement et de l'utilisation des systèmes et services reposant sur l'IA dans le domaine militaire ;

g) Élargir le champ d'application des documents et entités existants de l'Organisation des Nations Unies pour ce qui concerne les recommandations éthiques

sur le développement et l'application de l'IA, afin d'y inclure des lignes directrices propres à la conduite des conflits.

Le recours aux systèmes d'IA dans le contexte des conflits armés internationaux exige que la communauté internationale mène une action multilatérale de grande envergure, afin de garantir une utilisation responsable de ces systèmes. Les Nations Unies devraient jouer un rôle de premier plan dans l'instauration d'un dialogue, l'établissement de normes et le renforcement des capacités de la communauté internationale afin de prévenir la fragmentation et d'assurer une gestion appropriée.

Singapour

[Original : anglais]
[11 avril 2025]

En tant que petit État, Singapour a toujours appuyé le système multilatéral fondé sur des règles et le rôle joué par l'Organisation des Nations Unies. Celle-ci est à la base du droit et des normes internationaux. Les institutions, les systèmes et les instruments multilatéraux sont essentiels à la survie de tous les États, en particulier des petits États.

Singapour estime que les capacités d'intelligence artificielle (IA) exploitées dans le domaine militaire, y compris les systèmes reposant sur l'IA, doivent être utilisées de manière responsable tout au long de leur cycle de vie et en conformité avec le droit international applicable, en particulier le droit international humanitaire.

L'IA peut apporter des avantages dans le domaine militaire en améliorant la précision et la perception de la situation, ce qui permettrait de réduire les dommages collatéraux infligés aux civils ou aux biens de caractère civil. Toutefois, en l'absence de dispositifs de gouvernance appropriés, elle peut aussi présenter des risques d'intensification de conflits et d'erreurs d'appréciation. À cet égard, Singapour estime qu'il est crucial que la communauté internationale se penche sur cette question.

Approche de Singapour en matière de gouvernance de l'intelligence artificielle dans le domaine militaire

L'un des objectifs clés de la Stratégie nationale de l'IA 2.0 de Singapour est de promouvoir un environnement de confiance qui assure la protection des utilisateurs et encourage l'innovation. À cette fin, plusieurs secteurs du Gouvernement, y compris celui de la défense, élaborent des dispositifs de gouvernance de l'IA qui permettront de tirer parti de cette technologie, tout en veillant à ce que les risques liés à son emploi soient réduits.

À l'issue de concertations avec des technologues de la défense, des spécialistes de la planification militaire, des experts en droit international et des spécialistes de l'élaboration de politiques, Singapour a arrêté des principes nationaux sur l'IA dans le domaine militaire. Rendus publics en 2021, ces principes portent sur quatre principaux sujets de préoccupation :

a) **Responsabilité**. Premièrement, le risque de voir naître un nouveau comportement propre à l'IA doit être pris en compte. Les systèmes d'IA doivent être conçus pour des usages clairement définis, et les développeurs comme les utilisateurs sont responsables des résultats générés par ces systèmes ;

b) **Fiabilité**. Deuxièmement, le risque d'erreurs ou d'inexactitudes dans les résultats produits par un système d'IA doit être pris en compte. Les systèmes d'IA doivent être mis à l'essai et garantis à un niveau correspondant à l'usage auquel ils sont destinés. Ils doivent être conçus de manière à réduire au minimum les biais involontaires et à produire des résultats cohérents ;

c) **Robustesse**. Troisièmement, il faut parer aux risques d'exploitation de l'IA par des acteurs malveillants. Les systèmes d'IA doivent être conçus en tenant compte des cybermenaces et des menaces adverses liées à l'IA. Afin de remédier à l'effet « boîte noire », il convient de veiller à ce que leur développement soit suffisamment documenté pour pouvoir être expliqué ;

d) **Sécurité**. Quatrièmement, nous devons nous concentrer sur le risque de défaillance de l'IA dans des contextes critiques du point de vue de la sécurité. Les systèmes d'IA doivent pouvoir être utilisés en toute sécurité, s'agissant non seulement des plateformes déployées, mais aussi des biens et du personnel environnants.

Les principes directeurs susmentionnés éclairent l'approche de gouvernance de Singapour en matière de développement, d'essai, d'entraînement et de déploiement de systèmes reposant sur l'IA et conçus à des fins militaires.

Initiatives internationales et régionales sur l'intelligence artificielle dans le domaine militaire

Singapour participe activement aux initiatives internationales portant sur la gouvernance de l'IA dans le domaine militaire. En 2023, elle a soutenu l'appel à l'action pour une intelligence artificielle responsable dans le domaine militaire et la Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles. En 2024, elle a coorganisé le Sommet sur l'intelligence artificielle responsable dans le domaine militaire à Séoul (République de Corée), où elle a souscrit au plan d'action issu du Sommet.

Singapour reconnaît également l'importance des initiatives régionales, car elles permettent de tenir des débats inclusifs et propres au contexte sur l'IA dans le domaine militaire. En 2024, elle a coorganisé des consultations régionales sur l'IA responsable dans le domaine militaire pour l'Asie, ce qui a offert aux pays de la région un espace d'échange sur les possibilités et les risques que présente l'IA dans ce secteur.

En février 2025, Singapour et d'autres États membres de l'Association des nations de l'Asie du Sud-Est (ASEAN) ont adopté une déclaration conjointe sur la coopération dans le domaine de l'IA dans le secteur de la défense à l'issue du séminaire-retraite des Ministres de la défense de l'ASEAN qui a eu lieu à Penang (Malaisie). Dans cette déclaration, les Ministres se sont engagés à promouvoir l'utilisation responsable de l'IA, à approfondir au niveau régional la compréhension et la connaissance des répercussions de l'IA dans le secteur de la défense par l'échange d'informations, et à mettre en commun les meilleures pratiques et les enseignements tirés entre les États membres de l'Association.

Prochaines étapes pour les débats sur l'intelligence artificielle et la paix et la sécurité internationales à l'Organisation des Nations Unies

Singapour estime que tout débat futur visant à consolider l'appui de la communauté internationale à la résolution se doit d'être ouvert et inclusif. À cet égard, elle est favorable à la création dans le cadre de l'ONU d'un groupe de travail à composition non limitée axé sur l'IA dans le domaine militaire. Si un tel groupe de travail est créé, il devrait adopter une approche multipartite faisant notamment participer des technologues, des planificateurs militaires, des experts en droit international et des spécialistes des politiques. Singapour réaffirme son engagement à collaborer avec tous les États Membres afin de promouvoir l'application responsable de l'IA dans le domaine militaire.

Suisse

[Original : anglais]
[11 avril 2025]

1. Possibilités et risques

L'intelligence artificielle (IA) transformera probablement de nombreux aspects des affaires militaires. Elle s'annonce comme une technologie qui pourra appuyer les tâches et les opérations militaires, par exemple en améliorant la fiabilité, l'efficacité, la précision, la sécurité et la robustesse. Ses principaux domaines d'application sont notamment la perception de la situation, la prise de décision, le renseignement, la surveillance et la reconnaissance, la logistique et les chaînes d'approvisionnement, la formation et la simulation, ainsi que le commandement et le contrôle. Cette multiplicité d'applications est rendue possible par l'analyse de vastes jeux de données, qui permet une prise de décision plus rapide et plus éclairée. Dans le domaine de la surveillance et de la reconnaissance, par exemple, l'IA peut analyser des images prises par des drones et des satellites et détecter les mouvements avec une rapidité supérieure à celle des analystes humains. Elle pourrait également appuyer la reconnaissance des cibles en traitant les données des capteurs afin d'aider à distinguer les forces amies des forces hostiles. Dans le domaine de la logistique, l'IA peut permettre d'optimiser les chaînes d'approvisionnement, de prédire les pannes d'équipement et de veiller à ce que les ressources arrivent au bon endroit au bon moment. En ce qui concerne l'aide à la décision, les simulations réalisées au moyen de l'IA peuvent fournir aux commandants des informations prédictives et des résultats potentiels pour guider la planification stratégique. Les systèmes de formation et de simulation reposant sur l'IA offrent des environnements réalistes et adaptatifs qui contribuent à améliorer l'entraînement des soldats. Enfin, l'IA peut appuyer le commandement et le contrôle en rationalisant le flux d'informations, en améliorant la prise de décision et en renforçant la coordination entre les unités. Les systèmes reposant sur l'IA peuvent également aider à détecter les menaces, à renforcer la cybersécurité, à appuyer le maintien de la paix, à vérifier les accords de maîtrise des armements et à désamorcer les conflits au moyen de systèmes de détection lointaine, d'analyses prédictives et de dispositifs de surveillance, ce qui favorise la stabilité et la sécurité. Cependant, bien que ces avancées puissent procurer des avantages aux forces armées, l'intégration de l'IA dans le domaine militaire soulève également plusieurs préoccupations majeures et présente de nombreux risques.

Lorsqu'elle est utilisée de manière responsable dans les conflits armés, l'IA peut contribuer à renforcer le respect du droit international humanitaire et la protection des civils et des biens de caractère civil, par exemple en améliorant l'estimation des risques ou en augmentant la précision du ciblage afin de réduire les dommages collatéraux. Néanmoins, diverses applications de l'IA dans le domaine militaire en période de conflit armé, particulièrement celles à haut risque, suscitent aussi de graves préoccupations sur les plans juridique, humanitaire, éthique, de la sécurité et de la stabilité stratégique, lesquelles appellent une réponse, par exemple :

- **Erreurs dans la sélection des cibles**. L'IA a la capacité d'identifier techniquement des objets ou des individus à partir de ses données d'entraînement, mais la compréhension du contexte et les jugements de valeur, indispensables pour se conformer au droit international, représentent un défi particulier, ce qui pourrait conduire à une erreur d'identification d'objets ou de personnes comme cibles militaires, et ainsi à des frappes illégales ou involontaires.
- **Risques d'escalade**. Dans une crise qui évolue rapidement, un outil d'aide à la décision qui fonctionne comme une « boîte noire » pourrait recommander une

action agressive sans offrir de justification claire. En l'absence d'explications, les commandants peuvent soit suivre aveuglément des orientations erronées, soit perdre un temps précieux à les remettre en question.

- **Interprétation erronée des intentions**. Un système d'IA qui évalue le risque associé aux actions de personnes ou d'objets peut soulever des préoccupations (juridiques et de sécurité), en particulier lorsque les évaluations sont fondées sur des modèles dérivés de comportements et de contextes passés, sans contrôle ni jugement humains adaptés au contexte. Par exemple, un système d'IA surveillant le comportement de l'adversaire peut, en raison de données erronées, classer à tort des mouvements de troupes ordinaires comme hostiles, ce qui peut entraîner une action préventive et une escalade involontaire.

Ces risques soulignent l'obligation de veiller au respect du droit international existant, en particulier du droit international humanitaire, mais aussi le besoin urgent de poursuivre le dialogue et l'étude de cette question afin de mieux comprendre les risques et les défis, les mesures éventuellement nécessaires et d'examiner la nécessité, la valeur ajoutée et la faisabilité de la mise en place d'autres structures de gouvernance normatives. Il peut s'agir d'une législation nationale, de l'élaboration de meilleures pratiques, de normes ou d'instruments internationaux, ou de l'établissement de lignes directrices opérationnelles.

2. Cadre juridique

Le développement et l'utilisation de l'IA, comme de toute autre technologie, ne se font pas dans un vide juridique. L'IA militaire doit être développée, déployée et utilisée dans le plein respect du droit international existant, en particulier de la Charte des Nations Unies, du droit international humanitaire et du droit international des droits humains et d'autres cadres juridiques applicables. Aucune technologie ne doit jamais remettre en cause la validité du droit international. Le droit international, en particulier la Charte des Nations Unies dans son intégralité, le droit international des droits humains et le droit international humanitaire, s'applique et doit être observé et respecté.

Il incombe aux États et aux parties à un conflit de respecter et de faire respecter le droit international humanitaire en toutes circonstances, y compris lorsqu'ils utilisent l'IA dans le cadre d'opérations militaires. L'IA militaire devrait donc être conçue pour renforcer le respect du droit international humanitaire et la protection des civils et des biens de caractère civil. On pourrait y parvenir, par exemple, en veillant à ce que les systèmes reposant sur l'IA accordent la priorité à l'exactitude, à la réduction des dommages et à l'application du principe de responsabilité, notamment par des processus rigoureux de sélection, de validation et de vérification des cibles. En outre, l'IA devrait être utilisée de manière à renforcer le respect de l'obligation de prendre toutes les précautions possibles dans les opérations militaires, notamment pour éviter ou à tout le moins réduire autant que faire se peut les dommages accidentels, en aidant les commandants à protéger les civils et les biens de caractère civil tout au long de la conduite des hostilités, par exemple, en améliorant l'estimation des risques.

L'un des principaux domaines d'action est de veiller à ce que l'IA militaire soit conçue avec des jeux de données qui permettent son utilisation dans le plein respect du droit international, et qu'elle soit entraînée à l'aide de ces jeux de données. Au-delà de la conduite des hostilités, l'IA militaire doit être conforme à toutes les règles et à tous les principes applicables du droit international humanitaire, si elle est utilisée pour accomplir d'autres tâches régies par ce droit, par exemple en ce qui concerne la détention et l'internement de personnes ou le maintien de l'ordre et les mesures de sécurité publique dans les territoires occupés.

Lors du développement et de l'utilisation de l'IA dans le domaine militaire, on court le risque d'intégrer dans la conception des systèmes ou dans les données d'entraînement des interprétations juridiques trop laxistes, telles que l'élargissement de la définition des cibles licites ou le relèvement des seuils de dommages collatéraux admissibles. Appliquées à grande échelle, de telles interprétations pourraient progressivement compromettre la finalité protectrice du droit international humanitaire et accroître considérablement les préjudices subis par les civils. Ce risque souligne l'importance de préserver l'intégrité des normes juridiques, qui doit rester un élément central dans la gouvernance, la conception et le déploiement de l'IA militaire à l'avenir.

3. Interprétations et principes

Sur la base du cadre juridique susmentionné et dans son prolongement, et eu égard aux enjeux humanitaires, éthiques, de sécurité et de stabilité stratégique, il convient d'approfondir les interprétations et principes suivants :

1. Responsabilité, obligation de rendre compte et intervention humaines

- **Responsabilité et obligation de rendre compte**. Il incombe aux États de garantir que les êtres humains conservent en permanence la responsabilité et l'obligation de rendre compte, conformément au droit international applicable, pour toutes les décisions relatives à l'IA dans le domaine militaire.
- **Contrôle et jugement humains adaptés au contexte**. Les décisions militaires d'une importance capitale, que ce soit en salle des conseils ou sur le champ de bataille, et particulièrement celles qui concernent l'emploi de la force, doivent toujours être prises sous le contrôle et le jugement humains adaptés au contexte. L'IA militaire peut aider à la prise de décision, mais ne saurait remplacer les considérations et les jugements juridiques et éthiques ni se voir accorder une autonomie décisionnelle. Les États ne doivent intégrer ces systèmes que dans une chaîne de commandement et de contrôle où les êtres humains sont capables de conserver leur jugement et d'exercer des niveaux de contrôle appropriés. Les biais involontaires doivent être corrigés dans la mesure du possible.

2. Fiabilité, prévisibilité ou explicabilité et robustesse

- **Fiabilité**. L'IA militaire doit être fiable afin de permettre d'éviter des conséquences imprévues ou des dysfonctionnements, en particulier si elle peut avoir des répercussions négatives ou causer des dommages aux civils et aux biens de caractère civil. Elle ne doit être utilisée que si ses effets et ses conséquences peuvent être raisonnablement anticipés.
- **Prévisibilité ou explicabilité**. Les personnes responsables du déploiement de systèmes fondés sur l'IA doivent être en mesure de prédire et d'expliquer les décisions qu'ils prennent, afin de comprendre et d'anticiper leur comportement.
- **Robustesse**. L'IA militaire doit en outre faire preuve de robustesse, tant sur le plan technique qu'opérationnel, pour garantir la sécurité et la sûreté lors de son déploiement et de son utilisation.

3. Atténuation des risques

- **Amélioration de la perception de la situation**. L'IA devrait être utilisée pour améliorer la connaissance de la situation sur le champ de bataille, notamment en détectant la présence de personnes civiles afin de réduire la probabilité de dommages.

- **Analyse prédictive**. Il convient de recourir à des modèles prédictifs fondés sur l'IA pour évaluer les risques, élaborer notamment des stratégies de désescalade des conflits et éviter des victimes civiles.
- **Intégration de garde-fous**. L'IA militaire devrait intégrer des mesures de sécurité visant à réduire au minimum les dommages et à permettre une intervention humaine appropriée en cas de défaillance des systèmes.

4. Prévention de toute nouvelle dynamique d'escalade

- **Stabilité**. L'IA militaire ne doit être conçue, déployée et utilisée que de manière à ne pas exacerber les tensions internationales ou à ne pas créer de nouvelles dynamiques d'escalade.
- **Maîtrise des armements**. L'IA pourrait contribuer à la maîtrise des armements et ne doit en aucun cas compromettre les normes et instruments existants en matière de non-prolifération, de maîtrise des armements et de désarmement, ni entraver le respect de ces normes, notamment celles relatives aux armes biologiques et nucléaires.
- **Gestion des crises**. L'IA militaire pourrait aider à désamorcer et à gérer des crises.

5. Gestion du cycle de vie des systèmes militaires reposant sur l'intelligence artificielle

L'utilisation responsable de l'IA à des fins militaires exige une approche globale, tenant compte des risques et couvrant tout le cycle de vie de l'IA dans le domaine militaire, ce qui comprend la conception, le développement, les essais, le déploiement, l'exploitation, la mise à jour et la mise hors service de ces systèmes. À chaque phase, les considérations juridiques, humanitaires, opérationnelles et techniques applicables doivent être systématiquement prises en compte. Cette approche fondée sur le cycle de vie revêt une importance cruciale pour les applications militaires à haut risque de l'IA, comme les armes autonomes, la sélection de cibles ou l'aide à la décision, susceptibles d'entraîner la mort, des blessures ou des dégâts matériels, et, plus généralement, pour toute situation où les décisions sont régies par le droit international humanitaire. Pour les systèmes présentant un risque moindre, comme les outils d'appui administratif ou les systèmes de planification logistique, la gestion du cycle de vie devrait s'effectuer sur la base d'une estimation des risques propre au contexte.

- Lors des phases de conception et de développement, il incombe aux États de veiller à ce que les systèmes soient entraînés avec des jeux de données de haute qualité, représentatifs et comportant un minimum de biais, afin de permettre qu'ils soient utilisés en pleine conformité avec le droit international et les normes et règles y afférentes, et de limiter autant que possible les biais indésirables.
- Durant les phases d'essai et d'évaluation, des procédures rigoureuses de validation et de contrôle doivent être suivies afin de confirmer la fiabilité, la conformité légale et la robustesse opérationnelle dans des conditions réalistes.
- Lors des phases de déploiement et d'utilisation opérationnelle, des garanties doivent être mises en place pour surveiller les performances du système, assurer un contrôle et un jugement humains adaptés au contexte et permettre une intervention humaine adéquate.
- Durant les phases de mise à jour et d'apprentissage, les États doivent mettre en place des protocoles stricts pour toute modification du système, notamment une

gestion des versions rigoureuse, une revalidation et des procédures formelles d'approbation.

- Pour la phase de retrait ou de mise hors service, des mesures doivent être établies pour désactiver ou archiver les systèmes en toute sécurité afin d'éviter les utilisations abusives, l'activation ou le redéploiement involontaires.

4. Gouvernance internationale

La Suisse souligne qu'il importe de disposer d'un processus inclusif et continu dans l'Organisation des Nations Unies afin d'approfondir la compréhension commune des avantages, des risques et des enjeux liés à l'IA dans le domaine militaire et d'élaborer des principes visant à promouvoir une utilisation responsable de cette technologie. Il est donc impératif de mobiliser tous les États Membres, les parties prenantes concernées ainsi que les représentants du monde scientifique et technologique, de la société civile et du monde universitaire, afin de garantir la légitimité, la compétence et une large adhésion. Les instances connexes des Nations Unies devraient être transparentes, convoquées régulièrement et alignées sur d'autres initiatives menées sur la question.

L'objectif primordial de tous les efforts de gouvernance internationale visant à promouvoir une utilisation responsable de l'IA dans le domaine militaire doit être d'assurer le respect du droit international, en particulier le droit international humanitaire. En outre, les préoccupations humanitaires et éthiques, la préservation de la stabilité et l'atténuation des risques de sécurité doivent être au centre de ces démarches. La mise en place de dispositifs de gouvernance efficaces, l'adoption de normes communes et un dialogue multilatéral continu devraient contribuer à prévenir l'escalade involontaire, à promouvoir la transparence et la confiance mutuelle, et à consolider le rôle du droit international en période de bouleversement technologique. En fondant la gouvernance de l'IA dans le domaine militaire sur ces principes, les États contribuent à un environnement de sécurité plus prévisible, plus résilient et plus pacifique.

Voici quelques exemples concrets de mesures envisageables :

- Promouvoir l'uniformisation des interprétations, des définitions, de la terminologie et de la portée en ce qui concerne l'IA dans le domaine militaire ;
- Recenser et mieux comprendre les possibilités et les préoccupations relatives aux dimensions humanitaire, juridique, de sécurité et éthique ;
- Envisager des mesures de transparence et de confiance ;
- Élaborer des principes, des normes, des pratiques exemplaires et d'autres recommandations ;
- Donner des directives pour leur application.

Ukraine

[Original : anglais]
[11 avril 2025]

L'Ukraine développe et applique activement l'intelligence artificielle (IA) dans divers domaines d'activité, y compris le domaine militaire. Elle comprend clairement tant le potentiel de cette technologie pour améliorer le bien-être humain et renforcer les capacités militaires que les risques importants liés à son usage abusif dans le domaine civil et, plus particulièrement, dans le secteur militaire. Ces risques sont singulièrement élevés dans le contexte de l'invasion massive, non provoquée et injustifiée de l'Ukraine par la Fédération de Russie, au cours de laquelle cette dernière

viole systématiquement les lois et coutumes de la guerre et le droit humanitaire international.

L'Ukraine apporte son appui et prend part aux initiatives visant à dégager un consensus mondial sur le développement, le déploiement et l'utilisation responsables de l'IA civile et militaire.

À ce jour, l'Ukraine a participé à diverses initiatives, dont les suivantes : elle a signé la Déclaration de Bletchley en 2023 ; elle figure parmi les États qui ont souscrit à la Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles, publiée lors du Sommet de 2023 sur l'intelligence artificielle responsable dans le domaine militaire, qui s'est tenu à La Haye ; elle a soutenu l'appel à l'action pour une intelligence artificielle responsable dans le domaine militaire convenu lors du Sommet de 2023 et le plan d'action pour une intelligence artificielle responsable dans le domaine militaire, adopté comme document final du Sommet de 2024 sur l'intelligence artificielle responsable dans le domaine militaire ; elle a adhéré à la Déclaration sur une intelligence artificielle inclusive et durable pour les peuples et la planète lors du Sommet de 2025 pour l'action sur l'intelligence artificielle qui s'est tenu à Paris ; et elle fait partie des coauteurs des trois résolutions que l'Assemblée générale a adoptées jusqu'à présent sur l'IA, y compris la résolution 79/239 intitulée « L'intelligence artificielle dans le domaine militaire et ses conséquences pour la paix et la sécurité internationales ».

L'Ukraine se tient prête à prendre une part active aux nouvelles initiatives mondiales visant à promouvoir le développement sûr, éthique et responsable de l'IA. Elle appuie en outre les débats sur l'IA sous ses divers aspects qui se déroulent au sein du système des Nations Unies, notamment au Conseil de sécurité.

Nation pacifique n'ayant aucune revendication territoriale contre tout autre pays et victime de l'agression militaire perpétrée par la Russie, dont elle ne reconnaît aucune des revendications de cette nature formulées à son égard, l'Ukraine développe et utilise l'IA militaire exclusivement afin de renforcer ses capacités de défense, exerçant ainsi le droit à l'autodéfense prévu par la Charte des Nations Unies.

Selon l'Ukraine, l'utilisation de l'IA dans le domaine militaire présente les principaux risques suivants pour la paix et la sécurité internationales :

- La concurrence en matière d'intégration de l'IA dans les systèmes de combat et d'armes pourrait déclencher un cycle nouveau et plus dangereux de la course mondiale aux armements, compromettre la réalisation des objectifs de développement durable et surtout favoriser l'émergence de systèmes d'armement entièrement autonomes fonctionnant sans intervention humaine ;
- À l'instar d'autres technologies numériques, et compte tenu de la menace grandissante des cyberattaques ainsi que de la complexification et de l'élargissement des champs d'application, l'IA intégrée aux systèmes militaires s'avère de plus en plus vulnérable aux cyber-interférences et à la cyber-manipulation de la part d'une entité malveillante qui cherche à détourner ces systèmes de leurs fonctions initiales et de leur capacité de ciblage sélectif ;
- La dépendance excessive à l'égard de l'IA pour la prise de décision pourrait entraîner une perte de contrôle humain sur les processus militaires critiques ;
- L'intégration hâtive d'une IA sous-développée dans les systèmes d'armes peut provoquer des effets indiscriminés et une hausse du nombre de victimes civiles, surtout si cette IA présente des capacités défectueuses d'identification de cibles ;
- Il n'existe actuellement aucun cadre multilatéral visant à contrôler la prolifération des armes qui intègrent l'IA ;

- L'emploi d'armes intégrant l'IA soulève de sérieuses préoccupations juridiques et éthiques lorsqu'il n'est pas conforme aux lois et coutumes de la guerre et au droit humanitaire international.

B. Union européenne

[Original : anglais]
[11 avril 2025]

L'Union européenne se félicite de l'occasion qui lui est donnée de présenter ses vues sur les possibilités et les difficultés que l'application de l'intelligence artificielle (IA) dans le domaine militaire présente pour la paix et la sécurité internationales, en application de la résolution 79/239, adoptée par l'Assemblée générale le 24 décembre 2024.

Tout d'abord, l'Union européenne tient à rappeler sa position de longue date selon laquelle l'utilisation de l'IA dans le domaine militaire doit être conforme au droit international, notamment à la Charte des Nations Unies, au droit international humanitaire et au droit international des droits humains.

De même, l'Union européenne souhaite rappeler une autre position de longue date, à savoir qu'il faut toujours maintenir le jugement et le contrôle humains sur l'emploi de la force. La responsabilité et l'obligation de rendre compte liées à l'usage de l'IA en contexte militaire doivent continuer d'incomber aux êtres humains, afin de garantir l'application responsable de cette technologie.

L'Union européenne reconnaît que l'intégration de l'IA aux systèmes militaires offre des possibilités, mais présente aussi des difficultés. Le développement de l'IA est si rapide qu'il n'est actuellement pas possible d'en prédire tous les avantages ou les risques.

À cet égard, l'Union européenne se félicite de l'attention que les Nations Unies accordent à cette question ainsi que des débats qui ont lieu dans les instances internationales compétentes. Elle apprécie tout particulièrement la poursuite du processus sur l'intelligence artificielle responsable dans le domaine militaire, qui a débuté aux Pays-Bas en 2023 avec le premier Sommet sur l'intelligence artificielle responsable dans le domaine militaire, suivi du Sommet organisé par la République de Corée en 2024. Elle se félicite notamment du prochain Sommet qui se tiendra en 2025 en Espagne, et remercie l'Espagne pour son organisation.

L'Union européenne note que l'appel à l'action de 2023 pour une intelligence artificielle responsable dans le domaine militaire et le plan d'action de 2024 pour une intelligence artificielle responsable dans le domaine militaire ont été approuvés par tous ses États membres. Elle estime que le concept du Sommet sur l'intelligence artificielle responsable dans le domaine militaire, qui prévoit des processus multipartites et inclusifs autour de la question, est une approche prometteuse. À cet égard, elle est consciente de l'importance que revêtent d'autres contributions récentes, telles que le Sommet international sur l'intelligence artificielle et le Sommet pour l'action sur l'intelligence artificielle organisé par la France les 10 et 11 février 2025. Elle estime également que les travaux menés dans le cadre de la Déclaration politique sur l'utilisation militaire responsable de l'intelligence et de l'autonomie artificielles sont une contribution précieuse au débat international plus large sur les conséquences de l'IA pour la paix et la sécurité internationales.

Tous les États membres de l'Union européenne ont signé les documents finaux du Sommet sur l'intelligence artificielle responsable dans le domaine militaire et la Déclaration politique, et l'Union européenne pense que ces textes sont complémentaires et essentiels pour approfondir la réflexion mondiale, renforcer la

gouvernance et trouver les solutions pratiques afin de garantir une utilisation responsable de l'IA militaire.

Elle reconnaît que l'application de l'IA dans le domaine militaire présente des avantages, notamment en ce qui concerne la vitesse, l'ampleur et la précision des opérations militaires. L'IA peut fournir un avantage tactique grâce à la gestion et au prétraitement de vastes jeux de données provenant des systèmes de surveillance et d'armes, des drones et des images satellites, ce qui permettrait aux opérateurs humains de prendre des décisions plus rapides et mieux éclairées. L'IA peut aider à réduire les coûts en optimisant la logistique ou la maintenance de l'équipement grâce à la maintenance prédictive. Elle peut aussi permettre d'effectuer des opérations militaires à plus grande distance et avec une plus grande précision dans des environnements incertains.

Parallèlement, l'augmentation de la rapidité et de l'ampleur des opérations favorisée par l'IA militaire est un atout, mais engendre également des difficultés. L'IA accélère l'exécution de la boucle Observation, Orientation, Décision, Action. L'augmentation de la vitesse et de l'ampleur des opérations peut donner lieu à des erreurs de perception dues à des incohérences entre les intentions militaires et les analyses produites par les systèmes reposant sur l'IA. L'IA pourrait donc contribuer involontairement à l'escalade. La vitesse rend difficile la réalisation de l'objectif de conserver le jugement et le contrôle humains sur l'emploi de la force.

Dans ce contexte, l'Union européenne souligne l'importance d'une coopération internationale visant à étudier les répercussions de l'IA dans le domaine militaire et à envisager les cadres de gouvernance possibles.

Annexe II

Replies received from international and regional organizations, the International Committee of the Red Cross, civil society, the scientific community and industry[1]

A. International and regional organizations

African Commission on Human and Peoples' Rights

[11 April 2025]

I. Introduction

The African Commission on Human and Peoples' Rights (the African Commission), as the premier treaty-based human and peoples' rights body of the African Union (AU), is entrusted with the mandate of promoting and protecting human and peoples' rights in Africa under the African Charter on Human and Peoples' Rights (African Charter). In the African Commission's study on Addressing Human Rights Issues in Conflict Situations, the African Commission's Focal Point who led the study observed that 'it is … in conflict and crisis situations that the most egregious violations and abuses of rights are perpetrated…With the changes in the nature of conflicts and the attendant heightened threat to human and peoples' rights, there is a greater need for the human rights system to pay increasing attention to and provide effective responses to the challenges that these new dynamics present to the protection and observance of rights.' In the current context, one of the major new dynamics that carries serious implications for peace and security and therefore human and peoples' rights relate to Artificial Intelligence (AI) and in particular its rapid development and use in the military domain.

During its 1214th meeting, the AU Peace and Security Council (PSC), in requesting the AU Commission to conduct a study to assess the adverse impact of AI on peace and security, underscored the necessity of ensuring African perspectives in shaping global AI governance frameworks. Against this background and having regard to its work on AI and other technologies and human and peoples' rights[2] and human rights in peace and security, the African Commission is pleased to share its views in response to the invitation of the Secretary-General for submission of inputs on AI in the military domain and its implications for international peace and security.[3]

II. AI in the military domain and peace and security

The development and use of AI technologies in the military domain particularly to automate military functions such as surveillance, targeting, and the deployment of

[1] In accordance with operative paragraph 8 of General Assembly resolution 79/239, the replies received from international and regional organizations, the International Committee of the Red Cross, civil society, the scientific community and industry are included in the original language received. The Secretary-General remains committed to multilingualism as a core value of the United Nations.

[2] Resolution ACHPR/Res. 473 (EXT.OS/ XXXI) 2021 on human and peoples' rights and artificial intelligence (AI), robotics and other new and emerging technologies in Africa, available at https://achpr.au.int/en/adopted-resolutions/473-resolution-need-undertake-study-humanand-peoples-rights-and-art.

[3] The Focal Point of the African Commission on its study on human and peoples' rights and AI, robotics and other technologies acknowledges with appreciation the contribution of Professor Thompson Chengeta, who is the consultant providing technical assistance in the development of the study, through the Centre for Human Rights, University of Pretoria.

lethal force have far reaching consequences for peace and security and hence for human and peoples' rights. The AU Continental AI Strategy, endorsed during the 44th Extraordinary Session of the Executive Council of the African Union, highlights AI governance and regulatory challenges, particularly in military applications, warning that AI could exacerbate conflicts through inaccurate predictions or deployment of autonomous weapon systems. Additionally, the framework raises concern about disinformation, misinformation, cybersecurity threats, and military risks.

From the perspective of the development and use of AI in the military domain, peace and security should not be seen just from the perspective only of what it means for stability of states and societies. Beyond its conception under the UN Charter and public international law associated with friendly relations of states, peace and security is also a fundamental right of all peoples. The African Charter thus stipulates that 'All peoples shall have the right to national and international peace and security. The principles of solidarity and friendly relations implicitly affirmed by the Charter of the United Nations and reaffirmed by that of the Organization of African Unity shall govern relations between States.'[4]

The framing of peace and security as a right of peoples compels states to assess and govern the development and deployment of AI technologies in the military domain through a human rights lens that prioritises the prevention of harm, suffering, and injustice. Together international law conception of peace and security, it places an affirmative duty on states to ensure that AI systems do not contribute to conflict, perpetuate structural inequalities, or violate the rights and dignity of individuals and communities. By embedding peace and security within the framework of human rights, states are not only accountable for avoiding direct acts of aggression, but also for proactively creating and maintaining environments in which human flourishing, security, and justice are protected from the potentially disruptive or harmful impacts of emerging military technologies.

The implication of AI in the military domain to peace and security, farmed comprehensively, thus goes beyond how it shapes the obligation of states for non-aggression. It also covers how algorithm-driven systems may dehumanise individuals, introduce bias, and lead to unaccountable or disproportionate harm. It raises critical questions about the erosion of human oversight, the potential for unlawful killings or violations of international humanitarian law, and the targeting of vulnerable or marginalised populations.

By transforming military capabilities, the application of AI in the military domain can also have implications for peace and security by heightening tendencies for engaging in hostilities. The resultant escalation of tension and violence will be inimical not only to stability and peace between and within states but also most importantly carries more adverse consequences for the development needs of the less developed parts of the world such as Africa. While AI may contribute to advancing the development needs of Africa, its development and use in the military domain can have devastating consequences for development detrimental in particular to the right to development enshrined in Article 22 of the African Charter.[5]

This link between peace and development is also central to the Sustainable Development Goals (SDGs), especially SDG 16, which promotes peace, justice, and strong institutions. Without peace and security, sustainable development cannot be achieved. Recognising this link is critical in the governance of military AI, as the

[4] Article 23(1) of the African Charter.

[5] All peoples shall have the right to their economic, social and cultural development with due regard to their freedom and identity and in the equal enjoyment of the common heritage of mankind.

militarisation of AI can aggravate instability, particularly in fragile regions, and undermine Africa's developmental aspirations. By reaffirming the interconnectedness of peace and development, the African Commission calls for a governance approach that upholds peace as both a human right and a developmental imperative.

III. The need for a human and peoples' rights-based regulation of the development and use of AI in the military domain

Given the ways in which the use of AI in the military domain transforms the conduct of hostilities and how the development of AI relies on the extraction of natural resources particularly critical minerals such as rare earth minerals, it is the submission of the African Commission that both the process of extraction of resources in the development of AI in the military domain and the use of AI in the military domain need to be in full compliance with human and peoples' rights standards and international law principles, including international humanitarian law.

First and foremost, it is of paramount significance that the development and use of AI in the military domain complies with the right to peace and security enshrined in Article 23 of the African Charter on Human and Peoples' Rights. As a right that is born out of the recognition of the inseparability of the enjoyment of other human rights states from peace and security, this right entails that the use of AI in the military domain should be consistent with the international law prohibition of the use of force enshrined in the UN Charter and the Constitutive Act of the African Union.

Second, the use of AI technologies in conflict settings need to ensure respect for applicable human and peoples' rights and international humanitarian law principles, including most notably needs to adhere to the principles of precaution, necessity, distinction, proportionality and legitimacy. These requirements apply irrespective of whether the context in which the use of AI in the military domain relates to international armed conflicts or non-international armed conflicts. As established in the African Commission's study,[6] parties to conflict are obliged to observe human rights standards where such conflicts do not meet the IHL threshold of armed conflict. As such, those who use AI technologies in conflict situations that do not meet the IHL threshold of armed conflict are legally obliged to respect and ensure respect for the human and peoples' rights standards established under treaty and customary international human rights law.

Third, the development of AI in the military domain and the use AI technologies in hostilities need to comply with the principle of transparency. This is fundamental because it is the basis for ensuring effective regulation of the development and use of AI in the military domain and for compliance with applicable human rights and international law standards. Additionally, transparency is critical for ensuring compliance with the obligation for respecting the dignity, privacy and data protection of individuals. The principle of transparency is also a pre-requisite for addressing some of the concerns that arise from use of AI in the military domain including bias (owing to the source and type of data used) and explainability. Transparency is also critical not only with the development of AI in the military domain but also with respect to the transfer of AI technologies in the military domain.

Fourth, from the perspective of human and peoples' rights and IHL, the other standard key to human rights and international law-based regulation of the development and use of AI concerns accountability. In the event of the occurrence of violations of human and peoples' rights standards or IHL principles from the development and use of AI in the military domain, there has to be both institutional and individual accountability. Accountability in this instance encompasses not only

[6] ACHPR, Addressing human rights issues in conflict situations, https://achpr.au.int/en/node/895.

the measures that are taken against perpetrators but also the remedial steps that need to be put in place for redressing victims.

Firth, building and sharing of technical knowhow critical to ensuring regulation by states is the other principle. Recent developments including the jamming of GPS systems affecting flights reported in Eastern DRC and the deployment by the Islamic State of West Africa of armed drones, highlight not only the need for effective regulation but also the need for developing the requisite infrastructure and technical capacity for ensuring effective regulation.

IV. The link between the development of AI in the military domain and Africa's natural resources and its implications for peace and security

The African Commission is also of the view that when discussing peace and security, stakeholders must be aware of the link between development of military AI, Africa's natural resources – particularly critical minerals – and the notion of peace and security. Article 21(1) of the African Charter on Human and Peoples' Rights affirms: "All peoples shall freely dispose of their wealth and natural resources. This right shall be exercised in the exclusive interest of the people. In no case shall a people be deprived of it."[7] Article 21(5) further provides that "States parties to the present Charter shall undertake to eliminate all forms of foreign economic exploitation particularly that practised by international monopolies so as to enable their peoples to fully benefit from the advantages derived from their national resources."[8]

This provision is particularly important in the context of military AI, which depends heavily on critical minerals such as cobalt, lithium, and rare earth elements – resources abundantly found in Africa. The 2924 Report of the Chairperson of the African Commission's Working Group on Extractive Industries, Environment and Human Rights Violations, stressed the "significance of critical minerals for new and emerging technologies" and highlighted that Africa has been burdened by a "resource curse phenomenon."[9] The report of the Chairperson noted that "extraction of minerals and other resources not only fuels but also at times becomes the site where contestation over whose control and use triggers conflicts. In some instances, this has created a vicious cycle of insecurity and violence, a condition that not only leads to major human and peoples' rights violations but also the perpetuation of a vacuum of effective governance and the concomitant exploitative, socially and environmentally costly extraction of the resources of the continent."[10]

Therefore, governance of military AI must not only ensure the legal use of force but also address the exploitative chains of extraction that power such technologies. This requires strict oversight, equitable benefit sharing, and regional solidarity to prevent Africa's resources from being used to fuel further conflict and inequality.

V. Conclusion

The African Commission is of the view that the development and use of AI in the military domain carries far reaching consequences for international peace and security in general and for less developed parts of the world such as in Africa that historically suffered violations and remain vulnerable to the adverse impacts of the development and use of AI in the military domain without robust and effective legal regime for such development and use in the military domain. The African

[7] Article 21(1) of the African Charter.
[8] Article 21(5) of the African Charter.
[9] African Commission's Working Group on Extractive Industries, Environment and Human Rights Violations (2024), https://achpr.au.int/en/intersession-activity-reports/extractive-industries-environment-and-human-rights-violations (accessed 08 April 2025).
[10] As above.

Commission affirms that the development and use of AI in the military domain needs to be regulated on the basis of international law, human and peoples' rights and international humanitarian law standards with particular regard to the development and peace and security interests and human and peoples' rights needs of less developed parts of the world.

More specifically, beyond and above the right to peace and security, the governance of AI in the military domain needs to ensure respect for applicable human and peoples' rights and international humanitarian law principles, including most notably needs to adhere to the principles of precaution, necessity, distinction, proportionality and legitimacy, the principles of transparency, accountability and redress for victims and the obligation to build and share technical knowhow necessary for enabling societies to avert the risks that the development and use of AI in the military domain carries for peace and security. Only by ensuring that the development and use of military AI are aligned with international legal standards including those relating to the right to peace and security, the right to development, the right to privacy and protection of personal data, the right to remedy and the responsibility for exercising human control, the right to and control over natural resources and by addressing the structural inequities underpinning global technological advancement, can states uphold their duties to their peoples and advance genuine peace, justice, and security in relation to the development and use of AI in the military domain.

B. International Committee of the Red Cross

[19 March 2024]

Summary

The full submission is available at: https://www.icrc.org/en/article/artificial-intelligence-military-domain-icrc-submits-recommendations-un-secretary-general.

The International Committee of the Red Cross (ICRC) welcomes the opportunity to submit its views for consideration by the United Nations Secretary-General, in accordance with resolution 79/239.

The recommendations that the ICRC makes in this submission are in line with its long-standing mandate and practice of promoting respect for and the development of IHL, including its application to new technologies of warfare. This submission is intended to support States in ensuring that military applications of AI comply with existing legal frameworks and, where necessary, identifying areas where additional legal, policy, or operational measures may be required.

1. Normative proposals: Reaffirming existing IHL as the starting point

The ICRC has consistently emphasized that, while IHL does not explicitly prohibit or regulate the use of AI in military applications, it does restrict its development and use, and places strict constraints on AI when it is integrated into weapon systems or used in some way to conduct warfare.[1]

Existing and emerging normative proposals on the military application of AI should build upon established international legal frameworks and mechanisms, including IHL. Where necessary, these frameworks can be reinforced through the development of additional legal instruments, operational guidance or policy measures to address specific risks or challenges posed by emerging technologies. The form and content of such measures may vary depending on the specific use case. The ICRC

[1] This has also been affirmed by States, including in the UN General Assembly with Resolution 79/239.

encourages the international community to engage in concrete discussions on particular applications of AI in the military domain and to prioritize consideration of those that pose the greatest risks to people affected by armed conflicts.

2. A Human-centred Approach to military AI

In line with the resolution, the ICRC advocates for a human-centred approach to the development and use of AI in armed conflict.[2] This approach has at least two key dimensions: first, ensuring a focus on the humans who may be affected by the use of AI; and second, emphasizing the obligations and responsibilities of the humans using or ordering the use of AI in military operations.

Despite the growing development of AI-related technologies in the military domain, IHL requires individuals to make legal determinations. Humans must, for instance, determine the lawfulness of attacks that they plan, decide upon or execute, and they remain accountable for those determinations. The ICRC considers that human judgement is crucial for reducing humanitarian risks, addressing ethical concerns and ensuring compliance with IHL. Accordingly, while certain technical tasks may be carried out by machine processes, it is not the system itself that must comply with the law, but the humans using it.[3]

This does not mean that commanders and combatants cannot or should not use tools, including AI-decision-support systems. However, these tools must only be designed and used to support, rather than hinder or replace, human decision-making.[4] Further, States and parties to armed conflicts must ensure that human control and judgement are preserved in decisions that pose risks to the life and dignity of people affected by armed conflict. This is essential for ensuring respect for applicable laws, including IHL, and upholding ethical standards.[5]

3. Specific Applications of ai in the military domain

The ICRC has identified three specific applications of AI in the military domain that pose particularly significant risks to those affected by armed conflict:

1. AI in Autonomous Weapon Systems

Resolution 79/239 acknowledges the increasing integration of AI into weapons and weapon systems, a development that raises significant legal and humanitarian concerns. The integration of AI, particularly machine learning (ML) techniques, into autonomous weapon systems (AWS) exacerbates existing challenges posed by AWS in ensuring compliance with IHL. In particular, it increases difficulties for human users to understand, predict, and control the system's functioning and effects.

Users of AWS must be able to, with a reasonable degree of certainty, predict the effects of that weapon in order to determine whether it can be directed at a specific military objective, and take steps to limit those predicted effects, as required by IHL. This entails the ability to understand the functioning of the AWS: the nature and functioning of its sensors, the definition of its target profile and the potential effects in the circumstances of use, including any risk of error or malfunction. This is

[2] ICRC, AI and machine learning in armed conflict: A human-centred approach, 2019 (updated in 2021).

[3] ICRC, International Humanitarian Law and the Challenges of Contemporary Armed Conflicts: Building a Culture of Compliance for IHL to Protect Humanity in Today's and Future Conflicts (IHL Challenges Report), 2024, p. 61.

[4] *Ibid*.; ICRC, IHL Challenges Report – Chapter 2: Contemporary and future challenges in the conduct of hostilities, 2019, p. 32.

[5] ICRC, Decisions, Decisions, Decisions: computation and Artificial Intelligence in military decision-making, ICRC, 2024, p. 8.

particularly relevant for AWS that function in opaque ways (the "black box" challenge), such as AWS relying on AI techniques, which prevent the human user from being able to understand, predict or explain the system's output. This impossibility effectively results in a lack of control over the weapon's effects, rendering it indiscriminate by nature.

In this regard, we reiterate the joint call made by the ICRC President, with the UN secretary-general,[6] for new, legally binding rules prohibiting certain AWS and constraining the use of others.[7] In particular, we recommend a prohibition on

- unpredictable autonomous weapons – those that, due to their design or the circumstances and manner of use, do not allow a human user to understand, explain or predict the system's functioning and effects;
- autonomous weapons designed or used to target humans directly. This is required because of the significant risk of IHL violations and the unacceptability of anti-personnel autonomous weapons from an ethical perspective.[8]

The ICRC supports all efforts by States to urgently adopt a legally binding instrument to regulate AWS, in whichever forum they choose.[9] The integration of AI into AWS should also be considered when discussing normative proposals on military applications of AI. Doing so is essential to ensure a consistent and comprehensive approach to the regulation of military AI, to avoid normative gaps, and to effectively address the serious legal, ethical, and humanitarian risks that are exacerbated by the integration of AI into AWS. In this regard, the ICRC considers it important that binding prohibitions and restrictions on AWS, including AWS that incorporate AI, are integrated into broader discussions on the governance of military AI.

2. *AI in Military Decision-Making*

AI decision-support systems (AI-DSS) are computerised tools that bring together data sources – such as satellite imagery, sensor data, social media feeds or mobile phone signals – and draw on them to present analyses, recommendations and predictions to decision makers.

The use of AI-DSS raise concerns related to system functioning, data quality, and human-machine interaction. These systems risk increasing the rate of unforeseen errors, perpetuating problematic biases – particularly those based on age, gender, ethnicity, or disability, and making it difficult for the users to understand how and why the system generates its output from a given input.

Generally, AI-based systems will perform better when given well-defined goals and access to representative and high-quality data. However, armed conflict environments are marked by uncertainty, volatility, and deliberate deception techniques by adversaries, which makes it extremely difficult to obtain reliable or transferable data. Even where good data exists, it may not reflect the specific operational or humanitarian dynamics of a particular context.[10] Moreover, for AI systems that rely on training data, the utility of those data can rapidly diminish once

---

[6] ICRC, Joint call by the United Nations Secretary-General and the President of the International Committee of the Red Cross for States to establish new prohibitions and restrictions on Autonomous Weapon Systems, 2023.

[7] ICRC, ICRC Submission on AWS to the UN Secretary-General, 2024, p. 6.

[8] *Ibid*.

[9] *Ibid*.

[10] ICRC, IHL Challenges Report, 2024, pp. 64–65; ICRC, AI and machine learning in armed conflict: A human-centred approach, 2019 (updated in 2021); ICRC, Decisions, Decisions, Decisions: Computation and Artificial Intelligence in Military Decision-Making, ICRC, 2024, pp. 31 and 54.

a conflict begins. Parties to armed conflicts will continuously seek to maintain the initiative and operate in a manner that is not anticipated by their adversary, adapting their strategies and tactics accordingly. This can fundamentally alter the environment in which the system was expected to operate, making the original data no longer representative of the new operational conditions. In such cases, the system's outputs may become unreliable, and the AI model may require re-evaluation or retraining in order to remain fit for purpose.

Human interaction with these systems raises further concerns, such as "automation bias" – a propensity to rely on machine outputs even when other available information may call those outputs into question – which is particularly pronounced in high-pressure or stressful environments like in armed conflicts.[11] Taken together, these factors can hamper a user's ability to scrutinize the information available. The practical consequence might be, for instance, that someone plans, decides upon or launches an attack based solely on an AI-DSS's output, thereby effectively serving as a human rubber stamp rather than assessing the lawfulness of the attack by considering all the information reasonably available including the AI-DSS output.[12]

On the positive side, the careful use of AI-based systems may facilitate quicker and more comprehensive information analysis, which can support decisions in a way that enhances IHL compliance and minimizes risks for civilians. In the context of urban warfare in particular, the ICRC has recommended that online open-source repositories should be used to gather information about the presence of civilians and civilian objects.[13] Importantly, IHL imposes obligations to take constant care to spare the civilian population and to take all feasible precautions in attack. Therefore, in developing and using AI-DSS, armed forces should be considering not only how such tools can assist them to achieve military objectives with less civilian harm, but also how they might be designed and used specifically to protect civilians. However, the important point is that these computer outputs can inform but must not displace the need for legal determinations.

Beyond targeting decisions, militaries are also exploring the use of AI to support other operations traditionally carried out by humans, including detention operations. While technology deployed responsibly and with robust human oversight can contribute to IHL compliance, it also carries risks including bias, lack of transparency, and faulty programming and analysis, all of which can undermine compliance with IHL.[14]

To support efforts by States and other actors to ensure that military uses of AI-DSS remain consistent with IHL and humanitarian principles, the ICRC has formulated a non-exhaustive set of preliminary recommendations relating to the development and use of AI-DSS in armed conflict. They focus on 1) ensuring human control and judgement; 2) system design requirements; 3) testing, evaluation, verification and validation; 4) legal reviews; 5) operational constraints on use; 6) user training; 7) after-action reviews; and 8) accountability, among others. The recommendations are annexed to the full version of this submission.

[11] ICRC and the Geneva Academy, Artificial Intelligence and Related Technologies in Military Decision-Making on the Use of Force in Armed Conflicts: Current Developments and Potential Implications, ICRC, 2024, p. 17.

[12] ICRC, IHL Challenges Report, 2024, p. 65.

[13] *Ibid*., p. 66; ICRC, Reducing Civilian Harm in Urban Warfare: A Handbook for Armed Groups, 2023, p. 15.

[14] ICRC, IHL Challenges Report, 2024, p. 22.

3. AI in Information and Communications Technologies

AI is expected to change how actors defend against and conduct information and communications technology (ICT) activities, including in armed conflict. In particular, States have noted with concern that the use of AI and other emerging technologies in malicious ICT activities may further increase their scale and speed, as well as the harm they may cause.[15] For example, AI enables tools to identify and develop exploits for new vulnerabilities in software or networks, or to conduct harmful ICT activities autonomously, whether in offence or in defence. The ICRC is concerned that this could increase the risks of indiscriminate attacks, incidental civilian harm, including damage to critical civilian infrastructure, as well as the uncontrolled escalation of conflict, particularly in complex and interconnected digital environments.[16]

Similarly, information or psychological operations are not a new feature of armed conflicts; however, AI is changing how information is created and spread. AI-enabled systems, particularly generative AI, have been widely used to produce harmful content – text, audio, photos and video – which is increasingly difficult to distinguish from authentic, original content.[17] The ICRC is concerned about the consequences for civilians that might result from the creation and spread of such information through ICT, including information that contributes to or encourages violence, causes lasting psychological harm, undermines access to essential services or disrupts the operations of humanitarian organizations.

In light of these concerns, the ICRC underlines the importance of applying existing international law, including IHL, to the use of AI in ICT activities. The ICRC urges States to ensure that the development and use of AI-supported ICT activities respect the protections afforded to civilians and civilian infrastructure in armed conflict. Moreover, in light of the emergence of increasingly autonomous ICT capabilities, the ICRC further encourages States to address the serious challenges posed by these tools, particularly by considering whether existing international law, including IHL, provides sufficient safeguards against the harm such tools can cause, or whether additional limits are needed.

4. Conclusion

The ICRC is grateful for the opportunity to share its above views and recommendations on ways to address the challenges and concerns raised by AI for the secretary-general's consideration, and stands ready to contribute further to assist States in taking effective action to address the risks posed by AI applications in the military domain.

C. Civil society

Autonorms

[10 April 2025]

The following is the AutoNorms project's submission pursuant to Resolution 79/239 on "Artificial intelligence in the military domain and its implications for international peace and security" adopted by the United Nations General Assembly on 24 December 2024. The resolution requests the UN Secretary-General to seek

[15] 34th International Conference of the Red Cross and Red Crescent, Resolution 2 "Protecting civilians and other protected persons and objects against the potential human cost of ICT activities during armed conflict", 2024.

[16] ICRC, IHL Challenges Report, 2024, pp. 66–67.

[17] *Ibid*., pp. 58–59.

views, including those of Member States, civil society, the scientific community and industry, on "opportunities and challenges posed by the application of artificial intelligence in the military domain, **with specific focus on areas other than lethal autonomous weapons systems**". The AutoNorms team welcomes the opportunity for representatives of academia to submit their views on this important and timely topic.

The AutoNorms project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 852123). Led by Professor Ingvild Bode and hosted by the Center for War Studies at the University of Southern Denmark, the project examines how the integration of artificial intelligence (AI) technologies into weapon systems and military targeting shapes international norms governing the use of force.[1]

Introduction

Over the past 2-3 years, the international debate about applications of AI in the military domain has been characterized by two significant, near-simultaneous changes. First, there has been a **move away from its predominant focus on autonomous or AI technologies in weapon systems** towards considering AI technologies across a wider range of military decision-making tasks, especially in relation to targeting. To reflect his move, this submission focuses on the employment of **AI-based decision support systems (AI DSS),** or systems that are meant to be used as tools to directly or indirectly inform the complex process of use-of-force decision-making, for example, by analyzing large volumes of data, recognizing patterns within the data, predicting scenarios, or recommending potential courses of action to human decision makers.

Second, there has been a **growing emphasis on human-machine interaction** in the context of using AI in the military domain.[2] This emphasis results from the broad recognition that, even when humans are 'in' or 'on' the loop of targeting decision-making, they need to exercise a sufficient level of oversight, control, and agency over the targeting process. Human oversight is a governance principle featuring prominently across various international initiatives, including A/RES/79/239. However, **dynamics of human-machine interaction as part of the use of AI DSS both introduce new issues and solidify existing sets of challenges that require governance attention.** Our submission highlights these challenges and the need to ensure the exercise of human oversight and agency throughout the full targeting decision-making spectrum. It is structured in three parts, starting with explicating challenges of human-machine interaction, then commenting on the relative under-development of the international debate about AI DSS, and finally, sketching a way forward.

Challenges of human-machine interaction in the use of AI DSS

The use of AI DSS involves various dynamics of human-machine interaction because military personnel such as operators and intelligence analysts routinely and increasingly interact with a network of AI systems throughout the targeting process. These interactions involve multiple challenges **which have the potential to affect**

[1] The members of the AutoNorms team are Professor Ingvild Bode, Dr Hendrik Huelss, Dr Anna Nadibaidze, Dr Guangyu Qiao-Franco, and Dr Qiaochu Zhang. The AutoNorms project is based at the Center for War Studies, University of Southern Denmark, Odense, Denmark. For more information, please visit our website: www.autonorms.eu.

[2] Ingvild Bode and Anna Nadibaidze, "Symposium on Military AI and the Law of Armed Conflict: Human-Machine Interaction in the Military Domain and the Responsible AI Framework," *Opinio Juris*, April 4, 2024, https://opiniojuris.org/2024/04/04/symposium-on-military-ai-and-the-law-of-armed-conflict-human-machine-interaction-in-the-military-domain-and-the-responsible-ai-framework/.

the exercise of human agency, or humans' capacity to understand a system's functions and its effects in a relevant context; deliberate and decide upon suitable actions in a timely manner; and act in a way where responsibility is guaranteed.[3]

Dynamics of human-machine interaction result in **distributed agency between humans and AI systems, where they are not separated into two distinct entities but rather form part of a socio-technical system**.[4] As part of this system, both sides may influence each other in different ways, which then translate into various forms of distributed agency located along a spectrum. In some instances, dynamics of human-machine interaction will offer more opportunities for exercising human agency in targeting decisions. In other instances, however, the humans involved in use-of-force decision-making will be more constrained in their ability to exercise agency.

For example, **humans' ability to exercise agency might be limited by cognitive biases such as automation bias or anchoring bias**. Humans could over-trust AI DSS even when knowing that there might be malfunctions or unintended errors involved, risking an overreliance on algorithmic outputs without engaging in the critical deliberations and assessments that are needed to exercise human agency, especially in critical targeting decisions that might inflict death, destruction, and severe harm. Such biases are typically exacerbated by the increased speed of AI-assisted military decision-making, especially in contexts where there are high levels of pressure to act rapidly. They can also be exacerbated by AI DSS that are used for prescription or recommendations, because such systems restrict the options or courses of action available to human decision makers.

Moreover, given that AI DSS are likely to be employed not individually but rather as part of a network of systems, the increased complexity of interactions can result in situations where humans act upon some outputs suggested by AI DSS, but do not overall exercise a high quality of agency. Due to these and many other concerns related to interactions between humans and AI DSS, **there is a need to further investigate challenges of human-machine interaction that result in AI DSS not positively 'supporting' humans but rather undermining humans' ability to exercise agency**.[5]

The risks of not addressing challenges of distributed agency are substantial. First, situations where humans are restricted in their exercise of agency **raise questions about compliance with international humanitarian law**, which requires that humans be held accountable and legally responsible for violations of legal principles. Although humans remain officially in control of the selection and engagement of targets, there are concerns about the exact role played by humans in context of using AI DSS in practice.

Second, these concerns also extend to the risk of **negatively affecting moral agency and responsibility in warfare**. Challenges of human-machine interaction that result in distributed agency would allow humans to feel less morally responsible

[3] Anna Nadibaidze, Ingvild Bode, and Qiaochu Zhang, *AI in Military Decision Support Systems: A Review of Developments and Debates* (Odense: Center for War Studies, 2024), https://www.autonorms.eu/ai-in-military-decision-support-systems-a-review-of-developments-and-debates/.

[4] Ingvild Bode, *Human-Machine Interaction and Human Agency in the Military Domain*, Policy Brief No. 193 (Waterloo, ON: Centre for International Governance Innovation, 2025), https://www.cigionline.org/publications/human-machine-interaction-and-human-agency-in-the-military-domain/.

[5] Anna Nadibaidze, "Do AI Decision Support Systems 'Support' Humans in Military Decision-Making on the Use of Force?" *Opinio Juris*, November 29, 2024, https://opiniojuris.org/2024/11/29/do-ai-decision-support-systems-support-humans-in-military-decision-making-on-the-use-of-force/.

for decisions that could affect other people's lives. They also risk making the human role a nominal, 'box-checking' exercise which can *de facto* be compared with AI DSS playing an 'autonomous' role because the human role is substantially reduced.

Third, there are **security and operational risks related to distributed agency dynamics**, especially when they give too prominent roles to AI DSS and algorithmic outputs. AI systems often malfunction, are trained on biased sets of data which do not apply beyond the training context or specific contexts of use, as well as integrate assumptions that might not be strategically or operationally beneficial.

Various types of biases, issues of trust, uncertainties, targeting and military doctrines, political and societal contexts in which AI DSS are used – all these aspects **can lead to dynamics of distributed agency which limit the exercise of human agency and prioritize algorithmic outputs**. It is important to investigate these dynamics and ensure that distributed agency provides more opportunities than limitations to human decision makers in warfare.

Relative under-development of the international debate on AI DSS

Despite increasing reports about the use of AI DSS in recent and ongoing armed conflicts, and the significant challenges and risks they pose to the effective exercise of human agency, **the international debate on human-machine interaction in the use of AI DSS remains insufficiently developed**, particularly within intergovernmental UN settings. Current discussions on AI in the military domain, including those within the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (GGE on LAWS), have focused on the use of AI at the tail-end of the targeting process, specifically autonomy and AI in weapon systems. This narrow focus risks overlooking or failing to address critical normative, legal, ethical, security, and operational risks that can proliferate and compound throughout the entire targeting decision-making process.

An increasing, albeit still limited, number of stakeholders are raising this issue at international multistakeholder forums, such as the Summits on Responsible Artificial Intelligence in the Military Domain (REAIM). Some international non-governmental organisations and research institutes – such as the International Committee of the Red Cross (ICRC), the Stockholm International Peace Research Institute (SIPRI), the UN Institute for Disarmament Affairs (UNIDIR), and the Asser Institute – have initiated discussions on challenges posed by AI in the military domain, beyond the issue of autonomy in weapon systems. Despite this progress, there remains **a clear need to develop a more comprehensive and inclusive international multistakeholder debate to guide the responsible development and deployment of AI DSS in military contexts.**

Way forward

In closing, we sketch three ways intended to move the international debate about applications of AI in the military domain forward:

1. **Increase awareness for the implications of practices of designing, developing, and using AI DSS**. States and other stakeholders across industry, civil society, and academia engaged in the governance, development, and use of AI DSS for military targeting must consider the implications of their practices. These practices influence what counts as 'appropriate' ways of considering and employing AI DSS and thereby shape what becomes the accepted, requisite quality of human oversight and agency exercised over the whole process of use-of-force decision-making. To increase such awareness, the debate pursuant to A/RES/79/239 at the UNGA First Committee should centrally focus on the issue of AI DSS in the military domain.

2. **Consistently map both 'best' practices and 'problematic' practices associated with the design, development, and use of AI DSS**. To get a better sense of the direction that the design, development, and use of AI DSS take, states and other stakeholders need to closely map their own (and others') practices. While there have been some limited efforts to exchange potential best practices, we also need to be attentive to practices with potentially problematic effects. This should encompass practices exercised across the full life cycle of AI systems from development to use and post-use review. Mapping such practices would offer stakeholders a better overview of which practices may be beneficial, i.e., provide opportunities for the exercise of human agency, and which practices may be problematic, i.e., limit the exercise of human agency, and therefore assess the desirability of particular practices.

3. **Pursue the debate on AI DSS within a multistakeholder format**. States should work with diverse stakeholders – including academics across social sciences and technical disciplines, civil society representatives, and international organizations – to develop normative guidance and regulation, especially regarding the human role in military decision-making. Moreover, top-down processes towards governing AI DSS should be accompanied by a bottom-up, standard-setting process focused on establishing operational standards. Such an inclusive approach could strike a balance between national security and humanitarian concerns, while reinforcing the need to ensure that humans can exercise agency in use-of-force decisions.

Global Commission on Responsible Artificial Intelligence (AI) in the Military Domain

[11 April 2025]

1. Introduction

The Global Commission on Responsible Artificial Intelligence (AI) in the Military Domain (GC REAIM) welcomes the opportunity to contribute to the United Nations Secretary-General's report pursuant to resolution A/RES/79/239.

GC REAIM recognises that military applications of AI present both opportunities and challenges for global peace and security. Accordingly, the establishment of responsible and ethical governance – consistent with States' obligations under applicable international law – is essential. The global community must take proactive steps to ensure that military AI is developed and deployed in a manner that de-escalates rather than escalates conflicts; respects and enhances, rather than compromises, the sovereignty and territorial integrity of states; promotes rather than threatens the security and safety of civilians; constrains and supports rather than erodes the existing rules-based international order.

In line with GC REAIM's resolute commitment to advancing international governance efforts, this note outlines some of the – non-exhaustive – views expressed by GC REAIM Commissioners and Experts on the implications of AI in the military domain to peace and security. The views presented are general in nature and will be further elaborated in the forthcoming GC REAIM report. While the Commission plans to present substantive and actionable recommendations for stakeholders in September 2025, this note does not yet include concrete proposals. As discussions among Commissioners and Experts are still ongoing, it instead highlights some of the key opportunities, challenges, benefits and risks posed by AI in the military domain to peace and security.

2. Technological Foundations

GC REAIM holds that meaningful policy deliberations on AI in the military domain must be grounded in a shared, foundational understanding of the underlying technologies and their potential trajectories. The complexity of AI technologies often gives rise to misunderstandings, inflated expectations, or misguided applications. Consequently, it is imperative to demystify AI through formal and well-defined frameworks that distinguish between current capabilities and speculative future developments. To support this objective, GC REAIM is developing a taxonomy which seeks to map the full spectrum of AI applications across military and broader peace and security contexts. The taxonomy differentiates between the implications of AI in operational activities – such as warfighting and intelligence – and administrative activities – such as logistics and personnel training and helps identify the specific applications of AI that should be prioritised in governance deliberations.

In its approach to the creation of a taxonomy, GC REAIM highlights the need for and contributes to a concerted effort to clarify, standardise, and encourage the accurate use of technical language with different layers of abstraction for policymakers, experts, and the public, thereby enhancing transparency, mutual understanding, and public trust. GC REAIM also cautions against the uncritical multiplication or adoption of new terminologies in AI governance discourse, unless these are clearly defined; and to ensure such terms are not used to circumvent or obscure existing legal obligations. Precision and consistency in language are the basis of responsible AI governance.

3. Implications for Peace, Security, and Stability

GC REAIM recognises that the integration of AI into the military domain presents benefits as well as both foreseeable and unforeseeable risks to international peace and security. A balanced approach to the range of opportunities and challenges emerging throughout the AI life cycle lies at the core of GC REAIM's method and is essential for responsible AI governance.

AI in the military domain may contribute to international peace and security in several important ways. At the developmental stage, the advancement of military AI capabilities may act as a deterrent to violence, as the mere development and presence of advanced technologies by responsible actors can encourage restraint by aggressors. Military AI may enhance early warning systems, strengthening conflict prevention strategies, and supporting arms control verification through AI-driven tools that foster transparency, trust, and cooperation among states – fundamental elements in conflict prevention. AI can also bolster national security and defence by improving the precision, accuracy, and efficiency of intelligence analysis and situational awareness, enabling real-time threat detection, and facilitating more efficient counterterrorism operations through predictive analytics and autonomous systems. AI-powered systems can rapidly process vast amounts of complex data, enabling military forces to make timely, informed decisions that may prevent escalation and support conflict de-escalation efforts. These traits can also help improve targeting accuracy and precision, potentially reducing the risk of collateral damage or fratricide – attacks on one's own forces – and aiding compliance with International Humanitarian Law (IHL) to protect the security of protected persons, such as civilians and non-combatants, during armed conflict. Military AI may also reduce certain forms of human bias and enhance accountability by providing precise data, surveillance, and real-time monitoring, enabling clear attribution of actions to specific actors. In these ways, AI offers meaningful opportunities to reinforce adherence to international law and ethical standards, strengthening the normative foundation of the rules-based international order underpinning global peace and security.

AI in the military domain also presents a range of risks. As with the development of other general-purpose technologies, the development of AI in the military domain may accelerate arms races. AI technologies driven by the commercial market may be repurposed by militaries or soldiers in need or increase the access of violent non-state actors to AI-enabled military capabilities, which may intensify ongoing conflicts and contribute to broader instability. There are also concerns that states could employ AI technologies to suppress human rights, entrench internal repression, and destabilise both regional and global peace.

Concurrently, as with AI more broadly, the environmental consequences of military AI – such as the energy-intensive demands of AI systems, resource extraction, and ecological damage from AI-enabled military systems – could aggravate resource scarcity and environmental degradation, fuelling tensions and undermining long-term peace. However, given the impact militaries have on civilian technology development, efforts to reduce the environmental impact of AI in defence settings could have far-reaching beneficial consequences for all uses of AI. As such, considerations of environmental impacts should be a component of responsible AI governance in the military domain.

The large-scale data extraction required for AI development could intensify geopolitical rivalries, facilitate intrusive surveillance, and create distrust through opaque and exploitative data practices. Such deployment of military AI may perpetuate discrimination and exacerbate social divisions, undermining stability and ultimately international peace and security.

There are simultaneously significant concerns regarding the potential of integration of AI within the command, control, and communication (C3) structures of nuclear weapons. A number of Commissioners and Experts have emphasised that this is a red line that must not be crossed. The commitment of several nuclear-armed states to human decision-making surrounding the employment of nuclear weapons is therefore applauded. Further, the development of large-scale lethal autonomous weapon systems – such as swarms of anti-personnel devices – risks creating a new category of weapons of mass destruction, posing serious threats to global peace and security. Relatedly, AI may lower the barriers to creation and use of nuclear, chemical, or biological weapons by state or non-state actors, thus generating new challenges for arms control and non-proliferation regimes.

Beyond these strategic risks, AI may affect the character of war and lower the thresholds for armed conflict. By increasing the speed of armed escalation and driving changes in the capabilities of weapons systems, AI in the military domain may reduce states' confidence in their deterrent capabilities – particularly in the face of cyber infiltration risk – thus influencing how decision makers receive, process, and act on information. AI in the military domain could also exacerbate asymmetric warfare and violence by widening technological disparities that could increase the likelihood of force being used prematurely or disproportionately.

Operationally, inaccurate AI systems used for targeting can undermine the security of protected persons under IHL by increasing the risk of indiscriminate attacks, violations of proportionality, and failure to distinguish between combatants and civilians. Closely related to this is the risk of fratricide due to potential errors in target identification or decision-making, which can undermine operational effectiveness, escalate conflict, and erode trust within militaries and alliances. Finally, there are views that the use of certain AI systems in the military domain can create accountability gaps absent clear rules. By complicating the attribution of responsibility for unlawful actions, the deployment of AI in the military domain could undermine key principles of international law and state responsibility for internationally wrongful acts. This may complicate efforts to hold individuals or

states responsible for violations, leading to a reduced deterrent effect against unlawful conduct. Without avenues to hold actors legally responsible, the enforcement of international law weakens, potentially destabilising peace, encouraging impunity, and exacerbating global insecurity.

4. Decision-Making and Responsibility

GC REAIM acknowledges the ethical and legal challenges that arise from integrating AI into military decision-making which may have a direct impact on preservation of peace and security. The relationship between human judgment and machine outputs is complex and without measures to ensure lawful, responsible and effective development and deployment, there can be an erosion of accountability and increased risks of unintended harm. As AI systems become more sophisticated and integrated within military capabilities, it is plausible that algorithmic decisions may become more commonplace across global battlefields, introducing moral and legal challenges regarding human control, oversight and judgment in diverse contexts.

To address these risks, GC REAIM promotes the need for context-appropriate human judgement over specific uses, capabilities and decisions of AI in military applications. The GC REAIM report will list considerations and conditions that underpin and support human responsibility, judgment and means of adequately evaluating relevant actions and decisions. This could include the introduction of technical standards for explainability, as well as maintaining appropriate human oversight in targeting decisions, assessments of precautions, proportionality and distinction, and other critical operational choices. However, given that the very definition of autonomy in machines suggests the minimisation or removal of the human, ensuring human responsibility and accountability may require focusing on human decision-making at earlier stages of a system's life cycle, as the systems structure the behaviour of all who work with it. Human oversight is essential to uphold state obligations under applicable international law, in particular, IHL.

Military AI systems must be designed not only to support all individual and collective agents in the military domain to be effective in safely carrying out their lawful tasks, but also to do so responsibly and without compromising or undermining their status as moral human agents. GC REAIM suggests that military AI based socio-technical systems need to be explicitly and demonstrably designed to adequately attribute and apportion responsibilities and is determined to contribute to this process. For the security of protected persons, parties to armed conflicts should at all times be able to demonstrate that everything possible has been undertaken to create the conditions under which military personnel can effectively apply extant and widely shared principles and laws of armed conflict to their own situation, when using or relying upon AI components in the execution of their tasks.

5. Governance and Regulation

In light of both the opportunities and risks associated with military AI, GC REAIM supports a comprehensive governance framework that implements authentic international law. GC REAIM reiterates that existing legal regimes provide a solid foundation for regulating AI technologies. Governance must incorporate and account for procedural safeguards (due diligence and legal reviews, transparency of testing, evaluation, and validation, accreditation, and verification), substantive obligations drawn from various branches of international law, and soft law tools (military doctrines, national policies and strategies, norms and standards). In principle, all relevant international legal frameworks must be considered and applied. These include, but are not limited to, the following: (1) international law (*jus ad bellum*) which regulates when and how states use force, codifying a general prohibition on the use of force and exceptions such as in the case of self-defence, (2) international

humanitarian law (*jus in bello*) which governs conduct during armed conflict and ensures the security of protected persons, (3) international human rights law.

GC REAIM further emphasises the critical role of international, regional, and domestic institutions in implementing and enforcing these legal norms. Effective governance requires collaboration across these levels and the inclusion of both binding (hard law) and non-binding (soft law) instruments. Soft law mechanisms, such as codes of conduct and ethical principles, can complement existing treaties and facilitate rapid, flexible responses to technological developments.

To address the diverse range of challenges surrounding the integration of AI into the military domain, GC REAIM supports proactive risk-mitigation and confidence-building measures. While binding regimes are challenging for general purpose technologies, there may be opportunities for rigorous monitoring, verification, and enforcement mechanisms inspired by successful global arms control regimes. For example, Commissioners and Experts have discussed ideas such as an Autonomous Incidents Agreement to reduce the risks of miscalculation among AI-enabled autonomous systems, or a committee or consortium that could set guidelines and recommendations surrounding the testing and evaluation of AI systems, including generative AI. GC REAIM also suggests that states and industries should consider adopting human-centred safety-by-design principles, implement red-teaming practices throughout AI system life cycles, and maintain clear chains of accountability for all actors. Only through robust multilateral dialogue and inclusive multi-stakeholder cooperation can AI be effectively governed to enhance peace and security rather than exacerbate global instability.

GC REAIM acknowledges that the development of a comprehensive governance framework for military AI faces several key challenges. First, there is the challenge of diverse interests and perspectives, with states, private companies, and civil society holding varying and sometimes conflicting views on the regulation of military AI. Second, the sensitivity surrounding national security and defence poses a significant barrier, as many states are reluctant to subject their military technologies to international scrutiny or regulation due to legitimate security interests. Third, achieving meaningful and substantive inclusivity in discussions is often difficult, as key stakeholders may be excluded or marginalised in decision-making processes. Fourth, a trust deficit between states, international organisations, and the private sector complicates efforts to establish cooperative governance. Fifth, the presence of crosstalk, incommensurability, and discursive dissonance arises due to the diverse backgrounds and expertise of stakeholders, making consensus-building challenging. Finally, these obstacles are compounded by the lack of clear frameworks that address the complex ethical, legal, and technical issues at the nexus of AI and the military domain. In light of these challenges, the final GC REAIM report will offer strategies to navigate and overcome these barriers in developing a robust governance framework.

6. Conclusion

GC REAIM observes that the rapid advancement and deployment of AI technologies in military contexts poses opportunities, challenges, benefits and risks for global peace and security. Balancing these considerations must be met with a technologically sound, inclusive, principled, and legally grounded approach to governance.

A clear understanding of AI's technological foundations is necessary to properly address its role in modern warfare. Ethical and legal responsibility should remain human-centred, and governance frameworks must rely on the robust application of international law, supplemented by cooperative multilateral efforts and soft law

instruments when appropriate. In its formation and deliberations, GC REAIM has had the opportunity to reflect upon the conversations happening in broader governance processes, finding ways to effectively bridge gaps between disciplines and regional perspectives.

GC REAIM urges the United Nations and all State Parties to place these principles at the heart of global discussions on the implications of AI in the military and broader peace and security, for the present and future generations. Only through concerted international cooperation, guided by a shared commitment to human dignity, peace, and justice, can we ensure that the future of AI in the military domain is one that strengthens our common security.

InterAgency Institute

[11 April 2025]

The InterAgency Institute was established in December 2020 as a digital think tank, founded by expatriate and Global South women as a collective of researchers. It is in this condition that we address this submission on "opportunities and challenges posed to international peace and security by the application of artificial intelligence in the military domain, with specific focus on areas other than lethal autonomous weapons systems," following A/RES/79/239. With this, we seek to craft a complementary set of suggestions to develop the policy discussion in points where understanding that AI encompasses a wide array of data-processing techniques, and may be integrated into different types of warfare, in multiple parts of the organization, and at different levels.

The InterAgency Institute would like to point to overarching trends that fall within our areas of expertise, namely: (1) a focus on the global south, specially in how to prevent furthering the security gap; and (2) in how interagency cooperation in a time of greater mistrust may be leveraged to ensure the integration of AI in the military does not. Additionally, we make the point that Decision Support Systems (DSS) create analogous problems when compared to Autonomous Weapon Systems (AWS).

1. Addressing the security gap between the Global South and the Global North

The increasing technological intensity and digitalization of the battlefield are likely to increase the capacity gap between countries in the Global North & South. The "optimization of war" entails furthering this discrepancy, augmenting threats, and deteriorating the global security landscape. The wide range of AI-enabled solutions represents discrepant utility levels across tools.

While some tools require a low threshold (thus providing usually an equally low ceiling), the systems that pose the biggest military advantage require a high knowledge threshold to be implemented, therefore, will likely not be open source, and will only be available to entities with sufficient means to develop or acquire them. Given the experience in past decades on multilateral forums, it is important to recognize that interest in access to these technologies will play a role in the negotiations.

In the long term, the current trend of "technological sovereignty" (or more specifically of restricted technological access due to global inequalities) may be transformed to undermine such technological control, creating far-reaching implications of this new revolution in warfare, involving stakeholders that may be reluctant to shape modern discussions due to a lack of current development of these technologies in their ecosystem.

2. InterAgency cooperation in times of distrust

These issues call for interagency cooperation at both the strategic and operational levels. The lack of interagency cooperation might lead to threat escalation and the eroding foundations for peace and security. Interagency cooperation should focus on formalizing specific channels for communication between different States, developing strategies for AI implementation that will not damage diplomatic relations, and generating more transparency in the interactions between agencies and contractors. The participation of different branches of government at the UN-level discussions is pivotal for a whole-of-government perspective in the deliberations. Beyond interagency cooperation at the governmental level, the wide array of applications of military AI calls for different sets of Confidence Building Measures (CBMs).

Since AI may be integrated in different warfare types and at different levels, its applications for different contexts have different ethical implications and consequences. Therefore, a monolithic understanding of risks posed by AI in the military context and consequently a unique set of CBMs would be inadvisable. CBMs for AI use in the strategic level of cyberspace will not be the same as CBMs for AI use in the tactical level of aerial warfare. Therefore, thinking about CBMs for military AI as a monolith will lead to inaccurate and in some cases inapplicable measures, undermining its effectiveness.

There is a necessity for sharing best practices in the introduction of AI into military procedures. In this sense, a trade-off should be made, prioritizing best practices that contribute to strengthening the aforementioned points of interagency cooperation and CBMs, and other practices that fall within the larger umbrella of strengthening international peace and security. Sharing of best practices relating to cybersecurity and reliability of the technology could also take place, but they should give priority to CBMs that focus on integration of AI at the strategic level and in manners that avoid the escalation of threats.

3. Decision Support Systems

Target identification or recognition via AI-enabled Decision Support Systems (DSS) entail analogue problems to Autonomous Weapon Systems (AWS). Digital dehumanization, lowering the threshold of violence, and automation bias are byproducts of that process that may only be avoided by the creation of red lines prohibiting such systems that replicate those concerns.

This problem stems not only from AI, but from a wider trend. Other data processing techniques that involve deterministic sorting of data that is not adequately processed by human operators also generate these problems. This caveat should be made to understand that not only systems with AI-enabled technology in DSS pose these kinds of threats, but a wider array of data gathering/processing techniques.

Conclusions and recommendations

- Formal interagency bodies to interface with multilateral AI/military tech negotiations
- Funding and support for academic research in the Global South focused on military AI implications;
- Regular technical-diplomatic summits focused on transparency, shared definitions, and threat perception;
- Prioritize capacity-building initiatives for Global South actors;

- Red lines and confidence building measures could be tailored to the specific technology and operational context;
- The discussions on Autonomous Weapon Systems encapsulate worries around AI-enabled Decision Support Systems. The creation of red-lines for these systems could benefit from building upon recommendations of the GGE on LAWS;

International Committee for Robot Arms Control

[11 April 2025]

The International Committee for Robot Arms Control (ICRAC) values the opportunity to submit our views to the United Nations Secretary-General in response to Resolution A/RES/79/239 "Artificial intelligence in the military domain and its implications for international peace and security."

Founded in 2009, ICRAC is a civil society organization of experts in artificial intelligence, robotics, philosophy, international relations, human security, arms control, and international law. We are deeply concerned about the pressing dangers posed by AI in the military domain. As members of the Stop Killer Robots Campaign, ICRAC fully endorses their submission to this report, and wishes to provide further detail regarding the concerns raised by AI-enabled targeting.

Increasing investments in AI-based systems for military applications, specifically AI-enabled targeting, present new threats to peace and security and underscore the urgent need for effective governance. ICRAC identifies the following concerns in the case of AI-enabled targeting:

1. AI-enabled targeting systems are only as valid as the data and models that inform them. 'Training' data for targeting requires the classification of persons and associated objects (buildings, vehicles) or 'patterns of life' (activities) based on digital traces coded according to vaguely specified categories of threat, e.g. 'operatives' or 'affiliates' of groups designated as combatants. Often the boundary of the target group is itself poorly defined. Although this casts into question the validity of input data and associated models, there is little accountability and no transparency regarding the bases for target nominations or for target identification. AI-enabled systems thus threaten to undermine the Principle of Distinction, even as they claim to provide greater accuracy.

2. Human Rights Watch research indicates that in the case of IDF operations in Gaza, AI-enabled targeting tools rely on ongoing and systematic Israeli surveillance of all Palestinian residents of Gaza, including with data collected prior to the current hostilities in a manner that is incompatible with international human rights law.

3. The increasing reliance on profiling required by AI-enabled targeting furthers a shift from the recognition of persons and objects identified as legitimate targets by their observable disposition as an imminent military threat, to the 'discovery' of threats through mass surveillance, based on statistical speculation, suspicion and guilt by association.

4. The questionable reliability of prediction based on historical data when applied to dynamically unfolding situations in conflict raises further questions regarding the validity and legality of AI-enabled targeting.

5. The use of AI-enabled targeting to accelerate the scale and speed of target generation further undermines processes for validation of the output of targeting systems by humans, while greatly amplifying the potential for direct and collateral

civil harm, as well as diminishing the possibilities for de-escalation of conflict through means other than military action.

Justification for the adoption of AI-enabled targeting is based on the premise that acceleration of target generation is necessary for 'decision-advantage', but the relation between speed of targeting and effectiveness in overall military success, or longer-term political outcomes, is questionable at best. The 'need' for speed that justifies AI-enabled targeting is based on a circular logic, which perpetuates what has become an arms race to accelerate the automation of warfighting. *Accelerating the speed and scale of target generation* effectively *renders human judgment impossible or, de facto, meaningless.* The risks to peace and security – especially to human life and dignity – are greatest for operations outside of conventional or clearly defined battlespaces. Insofar as the use of AI-enabled targeting is shown to be contrary to international law, the mandate must be to *not* use AI in targeting.

In this regard, ICRAC notes that the above systems present challenges to compliance with various branches of international law such as international humanitarian law (IHL), *jus ad bellum* (UN law on prohibition of use of force), international human rights law (IHRL) and international environmental law. In the context of military AI's implications for peace and security, *jus ad bellum*, a framework that prohibits aggressive military actions and regulates the conditions under which states may lawfully resort to the use of force, is the most relevant. In the same manner IHRL is important in this context because it is designed to uphold human dignity, equality, and justice – values that form the foundation of peaceful and secure societies.

International Humanitarian Law and Youth Initiative

[11 April 2025]

Artificial intelligence (AI) has gained a universal recognition during the 1950s'. Technological emergence has assisted humans in almost all facets of their lives thereby making work easier and faster. Moreso, the rapid growth of Artificial intelligence in technological field enthralling commercial investors, law makers, defense intellectuals and international competitors can be evidential in theoretical premises of international security. The use of Artificial intelligence (AI) in modern warfare particularly in the In the Middle East and North Africa, Ukraine/Russian armed conflict which has resulted in the killings of thousands of innocent civilians with women and children being the most vulnerable. The emergence of AI is expected to be utilized in improving all sectors in our daily lives However, its Negative application in the military domain continues to create Humanitarian crisis between warring parties making it of regional and international concern. The war in Gaza is one of the deadliest and most destructive war in history with technology playing a central role in enabling mass slaughter and destruction ranging from supplying the dystopian systems used to automate the killings and bombing. [1] Following the October 7 2023, there have been extensive reports evidencing the Israeli occupation forces use of surveillance technology, artificial intelligence, and other digital tool to determine who, what and when to attack in Gaza trip. Thus, this violates the principles of international humanitarian law which emphasize the necessity of distinguishing those in active combat and not[2] and to take necessary precautions when conducting an attack to minimize civilian harm.

[1] Accessnow. (October 2024) Big Tech and the risk of genocide in Gaza: what are companies doing? Available at https://www.accessnow.org/gaza-genocide-big-tech/.

[2] Article 48 of Additional protocol I of the Geneva convention.

IHLYI in this paper, responding to the request of the UN Secretary-General pursuant to a resolution A/RES/79/239, adopted by the General assembly on 24 December 2024 on Artificial intelligence in the military domain and its implication to international peace and security therefore, it analyzes AI In modern warfare, its implication to international peace and security and the role of technological companies in armed conflict.

Artificial Intelligence in Modern Warfare: A Legal and Humanitarian Perspective

The rules of international humanitarian law do not explicitly address the use of modern technological tools and artificial intelligence (AI) during armed conflicts. However, its core principles – such as distinction, proportionality, and precaution – remain applicable and binding on all parties. These principles require the differentiation between military objectives and civilians, and oblige parties to take all feasible measures to avoid or minimize harm to civilian populations. In recent years, militaries have contracted private companies to develop autonomous weapons systems. However, the armed conflict in Gaza stands out as one of the most prominent cases where commercially developed AI models – originally created in countries like the United States – have been employed in actual combat operations, despite the fact that these systems were not initially designed to make life-or-death decisions.

This shift highlights a troubling rise in the militarization of technology without clear legal or ethical oversight. While some of these tools may enhance operational efficiency, their unregulated use poses serious risks of human rights violations, especially amid a lack of transparency about how these tools function, the origin of the data they rely on, and the accuracy of their outcomes[3].

One of the most pressing concerns recently raised is the deployment of digital military tools based on unreliable data or flawed algorithms. Some of these systems depend on mass surveillance of Gaza's[4] population, including the collection of personal data prior to the outbreak of hostilities. Such practices raise legal and ethical questions regarding their compatibility with international obligations to safeguard privacy and prevent the misuse of personal information for the purpose of direct targeting.

Among the tools reportedly in use is a system that tracks population movement through mobile phone data to monitor evacuations from certain areas. Another generates lists of structural targets to be hit militarily. A third tool classifies individuals based on levels of suspicion regarding their affiliation with armed groups, while a fourth seeks to determine the precise location of a target in order to carry out a strike at the opportune moment. These tools largely rely on data extracted from mobile devices – whether through cell tower location information or GPS[5]. However, from a technical perspective, such data is insufficiently precise to confirm an individual's presence at a specific location at a given time, particularly in conflict zones where individuals frequently change phones or numbers. Over-reliance on this technology may lead to fatal mistakes, especially when a mobile phone is used as a substitute for verifying a person's actual presence in a targeted area. Legally, the use of such systems without taking all feasible precautions to protect civilians constitutes

[3] Human Rights Watch, "Israel: AI-Powered Targeting Systems May Be Committing War Crimes in Gaza", 2024.

[4] Associated Press, "Documents Reveal Israel's Use of AI Tools in Targeting Gaza", Investigative Report, 2024.

[5] Human Rights Watch (2024). Questions and Answers: Israeli Military's Use of Digital Tools in Gaza Available at Questions and Answers: Israeli Military's Use of Digital Tools in Gaza | Human Rights Watch.

a clear violation of international humanitarian law – particularly Article 57[6] of Additional Protocol I to the Geneva Conventions, which obliges parties to take constant care to spare civilian lives during military operations.

Given this reality, urgent questions must be raised about the future of AI in warfare and the legislative and legal mechanisms needed to regulate it. Without proper oversight, these tools risk becoming instruments of systematic human rights abuses rather than technologies aimed at ensuring greater protection for those affected by war.

Implications of Artificial Intelligence on International Peace and Security

Armed conflicts in various regions around the world, such as Gaza, Lebanon, Syria, Ukraine, and Libya, have had catastrophic humanitarian and security consequences. These conflicts have led to the mass displacement of civilian populations, depriving thousands of people of their basic rights such as food, water, shelter, and healthcare. These individuals live in dire humanitarian conditions, with a significant increase in deaths due to famine, thirst, and diseases caused by contaminated water, in addition to exposure to harsh weather conditions without protection.

In this context, the increasing use of artificial intelligence and drones as weapons in conflicts, particularly by Israel in the Gaza Strip[7], stands out. Since October 2023, there has been a notable escalation in the use of "quadcopters" to carry out precise and targeted strikes against civilians. These drones are equipped with data analysis algorithms and offensive capabilities, enabling them to target individuals based on tracking their movements or mobile phone signals.

According to documented reports, this technology has led to the death of more than 1,000 Palestinians by May 2024, including a significant number of women and children. This constitutes a grave violation of international humanitarian law, particularly Articles 51 and 57 of Additional Protocol I to the Geneva Conventions, which prohibit attacks on civilians and obligate parties to the conflict to take all necessary precautions to avoid harming them.

The concerns are not limited to the use of artificial intelligence against individuals but extend to the misuse of data. Relying on mobile phone tracking technologies (either through GPS data or cell tower signals) to pinpoint individuals' locations presents serious risks. Recent studies have shown that these systems do not provide enough accuracy to reliably determine someone's location, especially in conflict zones where phones may be swapped or disconnected frequently. This means that relying on these methods without field verification can lead to erroneous decisions, resulting in unlawful killings.

In a well-known case, a Palestinian woman named "Silah" was killed while carrying a white flag and leading her family to safety. After stepping onto a main street, she was targeted by a small drone that shot her in the head. This incident, witnessed by those around her, serves as a stark example of the disastrous outcomes of unregulated use of technology on the battlefield[8].

In Libya, drones played a decisive role in the battles between conflicting parties, particularly as many of these drones, including Turkish and Chinese models, were operated using data analysis systems to target objectives. Some of these systems are

[6] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), Article 57.

[7] TRTWORLD (2024) Quadcopter strikes: 1000 Palestinians killed by Israeli drones in one year. Available at Quadcopter strikes: 1000 Palestinians killed with drones in a year.

[8] Gaza grandmother gunned down by Israeli sniper as child waved white flag," *Times Kuwait*, November 2024, https://timeskuwait.com/news/gaza-grandmother-gunned-down-by-israeli-sniper-as-child-waved-white-flag.

believed to rely on artificial intelligence techniques for targeting, without legal oversight. The use of these tools in urban areas like Tripoli and Sirte has led to the deaths of civilians and extensive damage to infrastructure[9].

All of these events indicate that integrating artificial intelligence into managing and directing armed conflicts without an internationally binding legal framework to regulate its use could open the door to widespread violations, especially if these systems are not subject to independent and transparent oversight to ensure compliance with international humanitarian law and human rights.

Roles of Companies Developing AI in Armed Conflicts

Through a rapid increase in artificial intelligence and computer services, U.S. tech corporations have discreetly given Israel the ability to monitor and kill many more militants in Gaza and Lebanon more quickly. However, the death toll among civilians has also skyrocketed, raising concerns that these instruments may be causing the deaths of innocent people. Israel's recent wars are a leading example of commercial AI models developed in the United States being used in active warfare, despite concerns that they were not originally designed to help decide who lives and who dies.

For years, militaries have hired private companies to create customized autonomous weapons. Numerous American software companies have backed Israel's battles in recent years, including Microsoft and the San Francisco-based startup OpenAI. Under "Project Nimbus," a $1.2 billion contract signed in 2021[10] when Israel first tried out its in-house AI-powered targeting systems, Google and Amazon offer cloud computing and artificial intelligence services to the Israeli military. The military has made use of Dell and Cisco data centers and server farms. Palantir Technologies, a Microsoft partner in U.S. defense contracts, has a "strategic partnership" that provides AI systems to support Israel's war efforts, while Red Hat, an independent IBM company, has also supplied cloud computing technologies to the Israeli military.

Furthermore, through a number of programs, Microsoft also supplies Israel's government with services that have allegedly been used to help the Israeli military, police, Israeli Prison Service (IPS), and illegal settlement operations. Over 10,000 Palestinians are being held by the IPS as of October 2024; half of them have been detained without being charged or having a trial date scheduled. At least 310 medical professionals, UN employees, women, and children are among the Palestinian prisoners from Gaza who are presently detained in prolonged, secret, and incommunicado detention, where they are subjected to torture, mistreatment, and sexual violence and abuse, according to the UN Human Rights Office.

Companies are under obligation to respect human rights within their scope of operations. Companies that directly aid the offender – for example, by offering financial, logistical, military, or intelligence support – may be held criminally responsible for a crime committed during an armed conflict. Companies and their managers or executives may be held accountable in certain situations even if they had no direct involvement in the crime or no intention of supporting it. As the Office of the High Commissioner on Human Rights (OHCHR) noted, companies "should treat

[9] France 24. (2021). "Have Killer Drones Been Deployed in Libya?". France 24. Retrieved from https://rb.gy/1m6k43.

[10] APNEWS (2025). As Israel uses US-made AI models in war, concerns arise about tech's role in who lives and who dies. Available at How US tech giants' AI is changing the face of warfare in Gaza and Lebanon | AP News.

this risk in the same manner as the risk of involvement in a serious crime, whether or not it is clear that they would be held legally liable[11]."

In light of the concerns raised in this submission and their implications for international peace and security, IHLYI urges states to:

1. **Refrain from the use of AI in military applications**: States should immediately halt the use of artificial intelligence in military activities and establish national regulations and laws to prevent its deployment in warfare.

2. **Work towards a global ban on the military use of AI**: States should actively pursue international agreements and frameworks to ban the use of AI in military contexts, ensuring that no country utilizes AI for warfare.

3. **Avoid the development of autonomous and AI-enabled weapon systems**: States should refrain from developing autonomous weapon systems or AI-powered weaponry that could be used to target humans, ensuring human oversight and decision-making in military actions.

4. **Ensure the protection of personal data**: States must guarantee that personal data is protected from misuse by military forces, law enforcement agencies, border control, and private contractors collaborating with these entities.

5. **Promote accountability in AI development**: Technology companies, researchers, engineers, and financial institutions should commit to not supporting the development or funding of AI technologies designed for military applications, advocating for responsible innovation in line with humanitarian principles.

Peace Movement Aotearoa and Stop Killer Robots Aotearoa New Zealand

[21 May 2024]

Peace Movement Aotearoa and Stop Killer Robots Aotearoa New Zealand welcome the opportunity to contribute our views to the UN Secretary-General's report on artificial intelligence (AI) in the military domain and its implications for international peace and security. Our submission briefly outlines our involvement in this issue, and has three sections summarising our position on: a) A new international instrument on military use of AI and autonomy in weapon systems is urgently needed; b) Key focuses of a new international instrument; and c) Scope of a new international instrument. The points below are based on discussions with our member and supporting groups about the content of this submission.

Introduction

Peace Movement Aotearoa is the national networking peace organisation in Aotearoa New Zealand, established in 1981 and registered as an Incorporated Society in 1982. Our purpose is networking and providing information and resources on peace, humanitarian disarmament, human rights and social issues; and we have extensive national networks of member and supporting groups and individuals. We are a founding member of the Stop Killer Robots campaign and coordinate the national Stop Killer Robots Aotearoa New Zealand (SKRANZ) campaign.

SKRANZ was launched in April 2013 to support the global campaign, with a specific national focus on urging New Zealand to take national action to prohibit the development, production and use of autonomous weapon systems; and to take

[11] Accessnow (2024) Big Tech and the risk of genocide in Gaza: what are companies doing? https://www.accessnow.org/gaza-genocide-big-tech/.

international action to support negotiations on a new treaty to prohibit autonomy in weapon systems. Since 2023 we have widened our focus to include military use of AI as its perils became increasingly obvious.

(a) A new international instrument on military use of AI and autonomy in weapon systems is urgently needed

As outlined in our submission for the UN Secretary-General's report on autonomous weapon systems (A/RES/78/241) last year, it has been clear for some years now that rapidly developing technological advances in the use of force and increasing autonomy in weapon systems pose an unprecedented threat both to humanity and to the foundations of international human rights and humanitarian law, which are based on respect for human life and dignity, protection of humanity in times of oppression and armed conflict, and human responsibility and accountability for harm.

The serious ethical, humanitarian, legal, and security concerns posed by these developments have been discussed for more than a decade within United Nations bodies – including the Human Rights Council, meetings related to the Convention on Certain Conventional Weapons and in the UN General Assembly – as well as in regional and national governmental and non-governmental forums.

Even as these discussions have taken place, some states have increasingly incorporated autonomy into military use of force in ways that have already resulted in gross violations of international law with disastrous consequences for civilian populations. It is apparent that the absence of specific international law on autonomy in weapon systems, and with differing interpretation by some states as to how existing law applies to new technological developments, the risk of proliferation of ever more dangerous and uncontrollable weapon systems is increasing rapidly.

The need for urgency for international action on this has been highlighted over the past eighteen months by, for example, Israel's use of AI-powered target suggestion systems in Gaza to make high explosive strikes on numerous targets possible in a short time frame, resulting in indiscriminate slaughter of civilians and systematic destruction of life-sustaining infrastructure. The reality of digital dehumanisation with catastrophic consequences is now very evident, as is the increasing tendency towards the development and use of autonomous weapon systems that will remove any remaining vestige of humanity from war.

We have noted with concern that states who brought forward A/RES/79/239 include states that have armed and supported Israel's genocidal attacks on Gaza, and where big data tech companies contributing data storage and AI capabilities to Israel's military systems are based.

Similarly, 'responsible AI in the military domain' (surely an oxymoron) is being promoted by states already developing their own AI targeting and autonomous weapon systems, as a way of undermining the push towards a binding instrument to prohibit these critical threats to international peace and security.

The US 'Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy' has highlighted for us the risk of horizontal proliferation of both military use of AI and autonomous weapon systems as states that do not have their own capability in this regard move from interoperability to integration with the states of armed forces that do or that are developing it. In the case of New Zealand, for example – as it seeks to be a 'combat capable force multiplier with enhanced lethality' [12] – this involves closer military integration with the armed forces of

[12] See, for example, the 2025 Defence Capability Plan released this week.

Australia and the US: New Zealand endorsed the US 'Political Declaration' early last year specifically to be compliant with US military doctrine.

These unfortunate developments underscore the urgent need for a new international instrument on military AI and autonomy in weapon systems to clarify and strengthen existing law. The instrument must include both prohibitions and regulations, as outlined below, and must include military use of AI in combat.

As emphasised in the UN Secretary-General's 2024 report on autonomous weapon systems[13], negotiations on a new instrument must begin without any further delay, in a multilateral forum where states can come together to work constructively, where the voices of those whose lives have already been impacted by military use of AI and increasing autonomy in weapon systems can be heard, and where UN agencies, the International Committee of the Red Cross (ICRC), and NGOs are active participants.

(b) Key focuses of a new international instrument

While much of the work around military use of AI and autonomous weapon systems has focused on the issue of meaningful human control over the use of force, it is our view that the key underlying ethical imperative is preventing human beings from being targeted or attacked by any system utilising digital code and/or sensors. A prohibition on military use of AI and autonomy in weapons systems that are designed or used to target human beings must be the starting point.

Meaningful human control over the use of force clearly has an ethical component, but it is also a practical and legal means to ensure accountability for any autonomy in weapon systems that breach the key dictates of humanitarian law.

(c) Scope of a new international instrument

It is our view that a new international instrument should include overarching rules to establish a framework for evaluating current and future technological developments, while promoting increased compliance with international human rights and humanitarian law.

Such overarching rules would prohibit autonomous weapon systems that are designed or used to target humans, and lay out specific obligations to ensure meaningful human control over other systems: for example, that the human operator/s understand the capabilities and limitations of the system, are able to fully evaluate the context in which the system will be used, and are making mindful firing decisions rather than assuming the technology is accurate – this would act to regulate autonomy in weapon systems. It would be useful to specify that decisions made by states on their assessment of new or altered weapon systems that incorporate autonomous features or functions must be transparent.

Furthermore, in the context of the UN Secretary-General's forthcoming report on AI in the military domain and in the light of the awful consequences of military use of AI in Gaza, the scope of a new international instrument must go beyond autonomous weapon systems. It is very clear that there is a spectrum of harmful military use of autonomy, ranging from target decision support systems (as some have described systems such as Lavender), data-based targeting systems, generation of target lists by algorithm or AI, sensor-based targeting systems, through to weapon systems that combine these elements and incorporate varying degrees of machine learning to make target selection decisions and attack autonomously.

We note the 2023 Joint Call by the UN Secretary-General and ICRC President stated *"The autonomous targeting of humans by machines is a moral line that we must*

[13] Lethal autonomous weapons systems: Report of the Secretary-General (A/79/88), 1 July 2024.

not cross"[14], yet that has already happened – a point reiterated in the UN Secretary-General's 2024 report[15].

It is therefore our view that a new instrument must cover military use of AI – including systems that automate significant decision-making in the use of force, such as target generation, force deployment, and engagement – as well as autonomous weapon systems.

Finally, although we have referred in this submission to military use of AI and autonomy in weapon systems, prohibitions and regulations in a new international instrument must also apply to all coercive agencies of the state, including those used for policing and internal security, for border control, in corrections facilities and in places of detention.

Ploughshares

[11 April 2024]

Project Ploughshares, a Canadian peace research institute, has for over a decade focused its advocacy and research on the military applications of emerging technologies, including artificial intelligence (AI) and autonomous weapons. As AI systems are rapidly advancing and being tested in contemporary conflict zones, international governance frameworks have struggled to keep pace. Meanwhile, intensifying geopolitical competition increases the likelihood that AI technologies will be deployed in complex, dynamic environments for which they are not suited – raising significant risks for civilians.

The wide-ranging use of AI in military applications demands urgent and coordinated international attention. We encourage the Secretary-General and member states to focus on three particularly pressing areas: the use of AI in decision-support systems related to the use of force, the dual-use nature of AI technologies, and the widening capacity gap among states engaging in multilateral discussions.

AI decision-support systems

One area that remains insufficiently addressed in current international discussions is the use of AI in military decision-making, especially decisions about the use of force. Of particular concern are AI-enabled targeting tools such as "Lavender" and "Gospel," reportedly used in in Gaza. These systems are classified as "decision support" because a human is technically required to approve target selections. However, there is little transparency regarding how these decisions are made, how frequently AI-generated recommendations are rejected, or whether human operators fully understand how the AI systems reach their conclusions.

In practice, these systems raise the risk of "rubber-stamping," in which human oversight becomes superficial, thereby undermining the principle of meaningful human control and increasing the likelihood of harm to civilians. The potential use of such AI systems in early-warning, surveillance, reconnaissance, and nuclear command-and-control systems further amplifies these concerns.

To mitigate these risks, states must work toward clear norms, regulations, and training requirements that enhance operator understanding, counter automation bias, and ensure genuine human engagement in decision-making processes.

[14] Joint call by the United Nations Secretary-General and the President of the International Committee of the Red Cross for States to establish new prohibitions and restrictions on Autonomous Weapon Systems, 5 October 2023.

[15] As at note 3.

Dual-use challenges

AI's dual-use nature – its applicability to both civilian and military domains – creates further governance complexity. Civilian-developed technologies can be repurposed for military use without appropriate testing or safeguards, increasing the risk of conflict escalation, misuse, and error. Additionally, the accessibility of certain AI tools means that nonstate armed groups may also gain access, potentially using them to target civilians and infrastructure.

We urge states to develop policy mechanisms, including export controls, technology impact assessments, and multistakeholder engagement, to account for dual-use risks and promote responsible innovation.

Capacity- and knowledge-building

Current multilateral discussions reveal stark capacity disparities among states, many of which do not have the resources or technical expertise to participate meaningfully in governance efforts. To ensure inclusive and equitable global engagement, we recommend that states collaborate with the UN Office for Disarmament Affairs to strengthen capacity-building initiatives.

The scientific and academic communities also have a role to play in supporting the development of accessible resources and training materials. International forums, such as the upcoming REAIM Summit in Spain, should include dedicated sessions for knowledge-sharing, especially to support representatives from under-resourced states.

Final thoughts

The international community is at a crossroads. The accelerating militarization of AI demands robust diplomatic responses. We can – and must – move from aspirational principles to concrete, enforceable frameworks, by employing political will, inclusive dialogue, and cross-sector collaboration.

AI-powered warfare is no longer a theoretical risk; it is a present reality. Whether this new era enhances global security or undermines it will depend on the steps states take now to strengthen governance, manage technological competition, and uphold international humanitarian norms.

Without timely, coordinated action, the risks of accidental escalation and unintended conflict will only increase.

Soka Gakkai International

[10 April 2025]

The Soka Gakkai International (SGI) welcomes the opportunity to share our views on the important issue of artificial intelligence (AI) in the military domain. As an NGO whose work is guided by Buddhist principles, we urge that the United Nations, its Member States and other stakeholders take into careful consideration the impact of AI in the military domain from a standpoint of upholding and respecting human dignity.

Introduction

AI in the military domain is rapidly evolving and transforming modern warfare and international peace and security. These systems are being used for various purposes, including surveillance, autonomous weapons, decision-making support, and logistics. With such wide-ranging applications, the integration of AI technologies in military systems poses significant challenges. To better ensure compliance with

international humanitarian law (IHL) and uphold protection for civilians and combatants alike there are several issues that we may consider.

Lack of transparency and accountability

- If an AI system were to make an error – such as identifying a target incorrectly – it could be difficult to pinpoint the cause of the error, "the black box problem". Was it a flaw in the data used to train the AI, an issue with the algorithm, or a problem in the operational context? Without transparency within these systems, assigning responsibility is difficult.
- International laws and treaties, such as the Geneva Conventions, were created before AI systems became commonplace in warfare. Without global norms and legal frameworks, there is no consistent approach to ensuring accountability for AI decisions made in warfare.
- With inadequate accountability mechanisms in place, AI could be used for military strategies that violate human rights, suppress civil liberties, or engage in unethical operations.

Speed of decision-making and risk of escalation

- The ability of a military force to make decisions and execute actions faster than its opponent is increasingly viewed as having a strategic advantage. However, the drive for speed can lead to unintended and costly consequences.
- Decisions made too quickly without proper analysis or consideration can lead to poor outcomes, including tactical blunders, strategic missteps, or ethical violations.
- Instead of diffusing a tense situation or negotiating, if combatants react too quickly it could provoke an even greater confrontation, further escalation and prolonged conflict resulting in more human suffering including amongst civilians.
- The acceleration of decision-making processes closes down the possibility of meaningful human control, the growing trend to automate decision-making threatens the ability to achieve human oversight which is essential to facilitate compliance with IHL.

Bias in AI in the military domain

- AI bias refers to the presence of systematic and unfair discrimination in AI systems, such as historical bias, where systems may reinforce harmful stereotypes, bias in data processing and algorithm development which can lead to making biased decisions and bias in how the systems are used.
- AI bias in the military domain is a significant concern, particularly as AI systems are increasingly being integrated into defense and security operations. The potential for AI bias to emerge in these areas can result in human rights implications, exacerbating existing inequalities and lead to deadly consequences for certain groups.
- AI heavily relies on vast amounts of high-quality and reliable data for decision-making. There are several potential violations when it comes to obtaining this data including issues around privacy and surveillance, challenges of bias also arise when dealing with incomplete and inaccurate data.
- When AI systems are biased, they not only perpetuate inequalities but also contribute to the digital dehumanization[16] of marginalized groups.

[16] Digital dehumanization is a process where humans are reduced to data, which is then used to make decisions and/or take actions that negatively affects their lives.

Proliferation

- Nations may rush to develop AI-based military technologies to outpace their adversaries, which could lead to a destabilizing arms race and increased global tensions.
- Without regulation autonomous weapons systems in particular, could proliferate globally, including amongst non-state actors which could increase crime nationally and regionally, exacerbating social inequalities, overwhelm resources and infrastructures of countries, as well as undermine social and national security.

Conclusion

The issue of AI within the military contexts is complex, and without regulation, it could lead to serious consequences for global peace and security. The desire to speed up decision-making processes within this context has yet to be proven as an effective way of resolving conflicts and achieving peace and security. Furthermore, you cannot divorce AI in the military and AI in civil uses, a failure to address AI in a military context could have widespread repercussions in all spheres of civil life including law enforcement, border control, education, housing and health care. Fundamentally, AI is here to stay, how we utilize it in the military and in our lives will shape the course of humanity. We have the possibility and the responsibility to decide how we want to use technology, knowledge, and the world's resources. To use it in a way that uplifts humanity or degrades it? This is an urgent question that requires moral, ethical and courageous leadership.

Stop Killer Robots

[11 April 2025]

The Stop Killer Robots campaign welcomes the opportunity to submit our views to the United Nations Secretary-General in response to Resolution A/RES/79/239.

Established in 2012, we are a coalition of more than 270 non-governmental organisations working across 70 countries.[1] We seek to counter threats to humanity and human dignity through the adoption of a new international treaty to prohibit and regulate autonomous weapons systems.[2] We support the development of legal and other norms that ensure meaningful human control over the use of force, counter digital dehumanisation, and reduce automated harm.[3]

Building an effective international response to emerging technologies

Autonomous weapons systems, 'AI in the military domain,' and trends and developments in increasingly automated decision-making and action in the use of force – as well as in our lives and societies more broadly – are all part of the same concerning picture:

The growing influence of computer processing and algorithmic thinking increasingly shapes our interactions in the world and the outcomes available to us. There are clear threats to peace, justice, dignity, human rights, equality, responsibility and accountability, and respect for law. We are getting closer to machine processes determining whom to kill.

[1] See www.stopkillerrobots.org/about-us and www.stopkillererobts.org/a-global-push/member-organisations.
[2] See https://www.stopkillerrobots.org/our-policies/.
[3] See www.stopkillerrobots.org/vision-and-values/.

To address these challenges effectively, a comprehensive and holistic response is needed from the international community.

Adopting a legally binding instrument on autonomous weapons systems will be one critical component: we must draw basic red lines for humanity against the automation of killing, which brings under jeopardy both international humanitarian law and international human rights law, in particular the presumption of innocence, the right to equality and non-discrimination, dignity, and wipes away contextual circumstances of the target(s) in question. The UN Secretary-General's comprehensive report last year reiterated his urgent call on states to negotiate a legally binding instrument to prohibit and regulate these systems by 2026.

But, a new international treaty on autonomous weapons systems alone may not be enough. States must also reach agreement on preventing and addressing grave harm from other uses of emerging technologies. A whole set of strong international rules are needed that stop the erosion of meaningful human control and the slide towards greater digital dehumanisation and automated harm, across international and domestic practice, in armed conflict and in civilian life.

'Military applications of AI' are already contributing to civilian harm

The risks of integrating AI into the use of force in armed conflict reach far beyond those to peace and security between states: a holistic consideration of peace and security that considers dimensions such as ethical, legal, and humanitarian issues must be taken into account in the UN Secretary-General's report under resolution 79/239.

We are already seeing grave threats to civilian protection and human rights and huge harm being caused by AI and automation in the use of force. This is arising from the quest for speed in warfare, the reduction of people to objects, and issues such as automation bias that Stop Killer Robots has raised the alarm about for years.

We have been horrified by reports of the use of AI-powered 'decision support systems' by Israel in Gaza, which suggest human targets to strike.[4] According to reports, human approval of these suggestions in vast volumes at high speed has been minimal – entailing digital dehumanisation, the erosion of meaningful human decision-making and control (including through automation bias), and directly contributing to massive and devastating harm to civilians in Gaza, alongside other tools.[5]

Many other states are developing and using such 'decision support systems', which raise concerns around international humanitarian law, human rights law, and ethics. So far there are few reports on how these are being deployed, with what constraints and with what impacts. Nevertheless, the push by many states to develop and integrate AI and autonomy into decision-making and the use of force is a huge concern. The further use in hostilities of these kinds of tools by any state in the unacceptable ways that we have seen in Gaza must be prevented. Stop Killer Robots struggles to see how such uses could meet the definition of the responsible application of AI in the military domain given in resolution 79/239.

Further risks to peace and security, rights, and human dignity

The quest for greater speed through AI and automation – towards the goal of increasing the tempo of conflict to a point beyond human cognition in the pursuit of a military and strategic edge – is an extremely dangerous one for international peace

[4] 'Lavender': The AI machine directing Israel's bombing spree in Gaza, +972 Magazine https://www.972mag.com/lavender-ai-israeli-army-gaza/.

[5] Questions and Answers: Israeli Military's Use of Digital Tools in Gaza, Human Rights Watch, https://www.hrw.org/news/2024/09/10/questions-and-answers-israeli-militarys-use-digital-tools-gaza.

and security. These risks are further to the impact 'AI in the military domain' is already having on civilian protection. Risks include unwanted escalation, lowered political thresholds to the use of force, and arms race dynamics.

Technologies that can contribute to target selection (such as threat detection tools) and remote biometric surveillance (such as facial recognition) have already had documented negative impacts on human rights such as the rights to privacy, equality and non-discrimination, freedom of expression and peaceful assembly, and the freedom of movement. In the case of facial recognition for identification (1:n), the technology is considered by many legal experts as wholly incompatible with international human rights law.

That AI systems inevitably encode and reproduce the biases of our societies – including racism, sexism and ableism – and that such bias cannot be eliminated, is also well established. The use of such systems to process people in the use of force will inevitably lead to disproportionate – and multiplied – impacts on already marginalised and minoritised people. Integrating automation and AI into decisions and actions in the use of force against people contributes to digital dehumanisation – the process where humans are reduced to data, which is then used to make decisions and/or take actions that negatively affects their lives.

The relationship with autonomous weapons systems

Stop Killer Robots notes that the UN Secretary-General's report will be on the "application of artificial intelligence in the military domain, with specific focus on areas other than lethal autonomous weapons systems." It is important nevertheless to highlight that various applications beyond the boundary of autonomous weapons systems are closely linked to them.

Firstly, such tools could be integrated as components of autonomous weapons systems now or in the future. For example, a 'decision support system' could be used as an autonomous targeting system, connected to a platform tasked to strike targets on the list generated, based on processing sensor data. Secondly, these tools are linked not only practically, but raise and are part of the same picture of concern. Strikes undertaken based on the nominal human approval of targets generated by a decision support system do not sit far from strikes undertaken with an autonomous weapons system.

It is therefore important that states consider these issues in dialogue: many of the rules and principles developed for autonomous weapons systems on keeping control and rejecting automated killing will need to be extended (with adaptations) to other tools; and, how the development of AI in the military domain more broadly will impact the direction and challenges posed by autonomous weapons systems will need consideration.

Recommendations

Technologies incorporating AI and automation into the use of force in armed conflict are currently being deployed without specific agreed rules; the principles various states have proposed and committed to so far have been too weak and vague to prevent civilian harm and risks to peace and security.

All developments in autonomy and AI in the use of force which threaten our safety, security, and humanity must be urgently and adequately addressed through strong regulation by the international community, with unacceptable uses prevented.

States must:

- Move with urgency to negotiate and adopt a new international treaty to prohibit and regulate autonomous weapons systems;

- In International discussions, critically and meaningfully engage with the implications and real-world consequences of current practice in the use of tools that fall under ‘AI in the military domain,’ including acknowledging and examining humanitarian harm;
- Fully consider the legal, ethical, humanitarian, and peace and security risks of further development and use of such systems, whatever the perceived ‘benefits’ may be
- Work urgently to prevent unacceptable uses of technology and trends in development, through committing to develop strong norms for meaningful human control and against digital dehumanisation:
  - This should take place domestically, regionally, and internationally.
  - It must involve a comprehensive and holistic international response, including a legally binding instrument prohibiting and regulating autonomous weapons systems alongside other measures.
  - It should include consideration and development of the other legal instruments necessary to preserve meaningful human control and to protect human dignity against AI in the use of force.

Stop Killer Robots Youth Network

[10 April 2025]

The Stop Killer Robots Youth Network welcomes the opportunity to submit recommendations for consideration by the United Nations Secretary-General in response to Resolution 79/239 “Artificial intelligence in the military domain and its implications for international peace and security” adopted by the General Assembly on 24 December 2024. As a global network of young people under age 30 in over 50 countries working to secure a future free of automated killing, we have consistently advocated for the creation of a new treaty on autonomous weapons systems (AWS) – in particular, we insist on a total prohibition of anti-personnel autonomous weapons as we wish to build a world without such dehumanising weapons. While youth will inevitably face the risks of new weapons technologies, we remain underrepresented in the decision-making process and are often sidelined in forums that shape our interests. As youth who have grown up in an increasingly digital world, we wish to create a future where technology is used to promote peace, justice, equality, and human rights, not perpetuate violence.

With escalating conflicts and the rapid deployment of new weapons technologies around the world, there is an urgent need to reinvest in international law as a measure to build trust and achieve sustainable peace and security. The application of artificial intelligence (AI) in the military domain presents numerous challenges that concern us as youth, including digital dehumanisation, the gamification of violence, and the further erosion of human control and involvement over the use of force.

Military AI & AI systems already in use

Artificial intelligence has been progressively implemented in the military domain over the past decade, however, due to the opacity of military activities and development, the wide public has not been aware of this issue until recently when the active uses of AI systems have been mediatized. We have seen and monitored the use of AI systems to support the targeting of both objects and people. Unfortunately, the use of such systems have not been able to alleviate civilian suffering, for example, in Gaza where one third of victims are children and where too many civilian infrastructures, including critical infrastructures such as humanitarian camps,

hospitals[1], and schools[2], have been either directly targeted or indirectly impacted by the hostilities.

There have been other concerning uses[3] of AI systems outside of the military which need to be considered as they might be implemented in the military domain, mainly predictive AI and facial recognition. Predictive AI technologies have been used in the police and judicial systems since the early 2010s and have been shown to be ineffective, incorrect, and subject to reinforcing discriminatory behavior.[4] If predictive AI were to be implemented in the military domain, it could lead to the increasing risk of civilians being targeted as they could be labeled as possible fighters or being indirect victims of military activities due to the multiplications of targets with predicted military advantages. Facial recognition technologies (FRTs) are also of concern as they are also unreliable especially when it comes to the identification of non-white males. Facial recognition-enabled targeting in military operations must be prohibited as those systems cannot comprehensively analyse every factor that makes military personnel or civilians a target or not.

Digital dehumanisation

One of the main concerns we have about the use of AI systems in the military domain is the proliferation and banalisation of **"Digital dehumanisation"**. We define digital dehumanisation as the process whereby humans are reduced to data, which is then used to make decisions and/or take actions that negatively affect their lives. This process deprives people of dignity, demeans individuals' humanity, and removes or replaces human involvement or responsibility through the use of automated decision-making in technology.[5] Additionally, the increased speed and scale of target production through military AI erodes moral restraints in war and lowers the impact and capacity of decisions from human operators[6], thus enabling the AI systems to make decisions without meaningful human control, which further dehumanises the decision-making process.

Relying on (Big) data leads to problems

We also believe that the use of (big) data in the military leads to multiple issues which need to be considered.

One of the primary issues is the challenge of data labeling – the process of categorizing and tagging data to train algorithms. Inaccurate or biased labeling can have far-reaching consequences, particularly in the context of distinguishing between

[1] World Health Organization (2025), 'oPt Emergency Situation Update'. https://www.emro.who.int/images/stories/Sitrep_57.pdf.

[2] Save the Children (2025), 'Education Under Attack In Gaza, With Nearly 90% Of School Buildings Damaged Or Destroyed'. https://www.savethechildren.net/blog/education-under-attack-gaza-nearly-90-school-buildings-damaged-or-destroyed.

[3] Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica (2016), 'Machine Bias'. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[4] Will Douglas Heaven, MIT Technology Review (2020), 'Predictive policing algorithms are racist. They need to be dismantled'. https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/.

[5] Automated Decision Research (2022), 'Autonomous weapons and digital dehumanisation'. https://automatedresearch.org/news/report/autonomous-weapons-and-digital-dehumanisation-a-short-explainer-paper/.

[6] Marta Bo and Jessica Dorsey, OpinioJuris (2024), 'Symposium on Military AI and the Law of Armed Conflict: The 'Need' for Speed – The Cost of Unregulated AI Decision-Support Systems to Civilians'. http://opiniojuris.org/2024/04/04/symposium-on-military-ai-and-the-law-of-armed-conflict-the-need-for-speed-the-cost-of-unregulated-ai-decision-support-systems-to-civilians/.

combatants and non-combatants in conflict zones. If the data used to train military AI systems is flawed or biased, it can lead to disastrous mistakes, such as the targeting of innocent civilians or misidentification of threats.

A critical issue when relying on big data is that the nature data itself is often broken and is incomplete. This means that the data used to train AI models can be incomplete, outdated, or unrepresentative of real-world situations. Such flaws in data can lead to systems that fail to generalize properly, resulting in inaccurate or incorrect predictions and decisions. For example, in combat situations, a lack of diversity in the data used to identify individuals could lead to inaccurate targeting, with devastating consequences. Important data might be missing or poorly represented, such as the exact location of civilians or combatants, which can lead to AI failing to make informed and balanced decisions. In a war scenario, a system trained with data from a specific past conflict may not be capable of handling a new, unpredictable situation. For instance, an AI system that has been fed data from one particular type of conflict might struggle to apply that data to a war with entirely different characteristics, resulting in errors in target identification or incorrect decision-making.

Another significant problem is that many AI systems operate as black boxes. This means that while these systems make decisions and predictions based on the data they process, the decision-making process is not transparent or easily understood. In military scenarios, where the consequences of decisions are extremely serious, the lack of transparency is particularly concerning. If an AI system makes an error, such as wrongly identifying a civilian as a combatant, the absence of clarity about how the system reached that conclusion makes it nearly impossible to understand the origin of the error. This makes accountability difficult, as we cannot determine why the system acted in a particular way. The lack of explanation regarding the decision-making processes of AI also makes it impossible to correct or adjust the system's behavior, potentially perpetuating errors without the ability to fix them effectively.

Linguistic and cultural bias embedded in data which is used to train AI systems can create security vulnerabilities and catastrophically misinterpret communications, behaviors, and intentions across diverse cultural contexts, potentially triggering lethal automated responses to misunderstood signals.[7] These systems risk automating and amplifying existing prejudices at unprecedented scale and speed with life or death consequences in conflict zones where cultural misunderstandings could rapidly escalate into devastating military actions causing dire consequences.

Accountability

The inclusion of AI systems in the command and decision-making chains will indubitably lead to a lack of accountability and liability for those relying on these systems to make decisions. It will create a sense of distance and lack of liability on the consequences of a decision which mean that decisions may be made without specific, consistent and thorough analysis of the lawfulness and humane characters of the decision. Then, if an action taken using AI systems violates IHL, the people involved in the implementation and those involved in the decision-making should be held accountable and the use of an AI system shall never exempt people from their responsibilities.

We recognize that military operations are bound by multiple bodies of law – national law, International Humanitarian Law (IHL) and International Human Rights

[7] Jimena Sofía Viveros Álvarez, Humanitarian Law & Policy (2024), 'The risks and inefficacies of AI systems in military targeting support'. https://blogs.icrc.org/law-and-policy/2024/09/04/the-risks-and-inefficacies-of-ai-systems-in-military-targeting-support/.

Law (IHRL) – which need to be respected and implemented in order for operations to be lawful. Unfortunately, rules of engagement and of targeting – and all the exceptions – cannot be fully understood and implemented by AI systems. Concepts like doubt, proportionality, and the balance between humanity and necessity are inherently human judgments that cannot be captured by an algorithm. Machines cannot be trusted to uphold these standards on their own. Therefore, it is critical that AI systems never act in a vacuum and that humans retain oversight and decision-making power at all times.

What the future might look like

While AI theoretically has the potential to enhance precision and efficiency in military operations, its integration into warfare raises significant concerns about the future of global security. Autonomous weapons systems, capable of making life-or-death decisions without human control, introduce ethical dilemmas and risks of unintended consequences. The use of AI in military technology is likely to aggravate the existing arms race, as nations compete to develop increasingly sophisticated AI systems, widening the power gap between technologically advanced countries and those less developed, leaving them vulnerable in terms of military readiness. The deployment of autonomous weapon systems and AI-driven tools makes conflict more unpredictable, scalable, and asymmetric, granting certain nations the ability to unleash devastating technologies that smaller states or non-state actors may not be able to counter. The proliferation of AI in the military sphere also raises the threat of terrorism, as organized actors could easily access advanced AI-powered systems. Moreover, the fast-paced, constantly evolving nature of AI development turns military strategies into a “cat and mouse” game, where advancements are met with equally rapid countermeasures. In light of these challenges, the future of military AI must be handled with extreme caution, emphasizing robust ethical frameworks, international regulations, and stringent human oversight to prevent these technologies from destabilizing global peace.

What we need

We call for the establishment of a meaningful legally binding instrument for the use of AI-driven systems in the military requires comprehensive integration of the technical sector alongside state actors, addressing the urgent need for standardized verification protocols and trust-building mechanisms between nations. Such an instrument should define clear autonomy thresholds that specify permissible levels of independence in target selection and engagement, mandate extensive documentation of algorithmic decision processes and testing methodologies and establish explicit red lines that cannot be crossed including prohibited deployment scenarios, target categories, and operational environments. This framework should apply consistently across developing and developed nations, incorporate independent verification bodies with appropriate technical expertise to conduct regular compliance audits, and establish enforcement mechanisms with meaningful consequences for violations, all while facilitating technical data sharing and research that builds confidence between stakeholders in this domain.

These systems present an unprecedented threat to global security and human rights, and the risks they pose to non-combatants are immense. It is crucial that it implements a robust framework of monitoring, accountability and oversight. Firstly, the states need to be bound by positive obligations to ensure the responsible use of AI in the military domain. Accountability is a fundamental aspect of this framework. We call for comprehensive mechanisms that oversee every stage of the AI system life cycle, from development and updates to transfers and research. States must ensure that any uses of AI systems are monitored, with clear reporting structures in place to

address incidents promptly. Furthermore, it is vital that human operators using these systems receive thorough training and guidance to make ethical decisions in the field. The principle of meaningful human control must remain central when it comes to the use of AI in the military domain to ensure that ultimate responsibility for any actions remains with human decision makers.

Unione degli Scienziati Per Il Disarmo

[6 April 2025]

Introduction

USPID (***U****nione degli* ***S****cienziati* ***P****er* ***I****l* ***D****isarmo, Union of Scientists for Disarmament*) is an association of concerned scientists – founded in 1983 and based in Italy – which promotes arms control and disarmament initiatives based on scientific understanding of risks posed by military applications of science and technology. USPID submits to the United Nations Secretary-General its views on "Artificial intelligence in the military domain, with specific focus on areas other than lethal autonomous weapons systems, and its implications for international peace and security", in accordance with the invitation formulated in operative paragraphs 7 and 8 of Resolution 79/239 adopted by the UN General Assembly on 24 December 2024.

Hazards for peace and security arising from AI military applications

USPID expresses its deep concern about new hazards for peace, international security, and the respect of International Humanitarian Law (IHL) which arise on account of the ongoing and accelerating military efforts to incorporate Artificial Intelligence (AI) into multiple facets of warfare. Major sources of these hazards have been identified in current limitations of our capability to understand, predict precisely, and control the behavior of AI systems developed by machine learning methods and their interactions with other human or artificial agents. Initially identified in connection with the operation of AI-enabled Autonomous Weapons Systems (AWS), these hazards are now spreading to AI systems supporting intelligence collection, the achievement of situational awareness, and human decision-making in warfare.

Exceptionally grave concerns are raised by proposals to integrate AI in Nuclear Command, Control, and Communication (NC3) and in adjacent systems supporting nuclear decisions, and to let AI perform tasks that might directly or indirectly affect nuclear decision-making. A significant case in point is the proposal to use AI technologies in nuclear early warning and decision-support systems, which is being advanced with the expectation that AI accuracy will reduce potential errors, and its processing speed will buy more time for nuclear decision makers. However, on account of the probabilistic nature of AI information processing, one cannot exclude the risk of AI perception leading to false positives of a nuclear attack or producing perniciously unreliable recommendations given the impossibility of ensuring that the underlying models are aligned with human values and the UN overarching goal of preventing and removing threats to peace. If such mistakes occur, no matter how infrequent, large-scale and even existential implications for humanity might ensue. Accordingly, it would be imperative to proceed with time-consuming verifications of AI responses in nuclear early warning. But these verifications would be hindered by the black-box nature of much AI information processing and by the reliance on mostly simulated data, eventually thwarting the expectation of buying more time for human decision makers.

Additional concerns are raised by proposals to exploit the rapid pace at which AI operates to speed up battlefield decision-making and targeting cycles. These proposals are fueled by the goal of gaining military advantage over potential adversaries.

However, fighting at machine speed jeopardizes both the effectiveness of human oversight on AI-enabled decision support systems and the fulfilment of ethical and legal roles that are attached to human oversight of warfare action. Indeed, overly tight temporal windows for decision-making hinder effective human control over IHL threats raised by machine suggestions. Human interventions which aim at preventing inadvertent conflict escalations prompted by fighting at machine speed are similarly hampered. In addition to this, excessive speed in human-machine interactions has been identified as a factor inducing automation biases on the battlefield, and potentially skewing human decision-making even in the absence of AI failures.

Further hazards arise in connection with inherent vulnerabilities of AI learning methods and systems. Malicious manipulation of input data might be exploited to induce classification mistakes by AI systems. Moreover, poisoning attacks corrupting learning datasets may impair learning processes and the accuracy of resulting AI systems. These risks are compounded by our current inability to fully align AI systems with human goals and values, potentially causing them to deviate from strategic objectives.

Recommended actions

Mindful of these and other emerging hazards posed by the rapid adoption of AI technologies and systems in the military domain, USPID recommends

- to integrate discussion of AI in NC3 into the Non-Proliferation Treaty framework and in dedicated high-level dialogues and forums such as the Summit on Responsible Artificial Intelligence in the Military Domain (REAIM);
- to develop sustained international dialogue, good practices, and confidence-building measures concerning new and emerging risks for peace and IHL respect raised by AI warfare applications;
- to support a comprehensive and detailed inquiry aimed at identifying actual and potential AI applications in the military domain, jointly with situations of use that pose serious threats to peace, international stability, and the respect of IHL;
- to consider and investigate the need to introduce international regulations or prohibitions for those AI military applications that pose serious threats to peace, international stability, and the respect of IHL.

Women’s International League for Peace and Freedom

[24 May 2024]

The Women’s International League for Peace and Freedom (WILPF) has opposed war and the development of technologies of violence since its founding in 1915. WILPF has consistently condemned military spending and militarism as detrimental to human life and wellbeing. Our concerns with artificial intelligence (AI) in the military domain and its implications for international peace and security are grounded within our wider opposition to weapons, war, and violence, as well as in our opposition to patriarchal, racist, and colonial power relations that are embedded within AI technology.

While there are many perils of the military use of AI; WILPF’s submission is focused on the following issues:

1. The need for human emotion, analysis, and judgement in relation to the use of force;

2. The existence of gender, racial, and other bias in AI technology and the implications for digital dehumanisation;

3. The impacts of military use of AI on privacy and personal data;

4. The environmental harms exacerbated by the military use of AI; and

5. The dangers of war profiteering and arms racing.

Due to the concerns raised in this submission and in other spaces, WILPF opposes the military use of AI. This technology, rather than placing limits on violence or harm, expands both. Governance is insufficient in the face of the profits and power the developers of these technologies seek.

In light of the concerns raised in WILPF's submission and the implications for international peace and security, WILPF urges states:

- To refrain from using AI in the military domain and to develop national laws and regulations to this end;
- To pursue a global prohibition on the military use of AI;
- To not develop autonomous weapon systems or AI-enabled weapon systems, including those that can be used to target human beings;
- To ensure protection of personal data from use by militaries, police, border enforcement, and private companies and contractors collaborating with these institutions;
- To uphold human rights and dignity online and offline; and
- To address the environmental harms generated by data centres, cloud computing, and AI by reducing the number of these centres and energy consumption and water use, which will include reducing the overall use of AI.

WILPF also urges:

- Technology companies, tech workers, scientists, engineers, academics and others involved in developing AI or robotics to pledge to never contribute to the development of AI technologies for military use;
- Financial institutions such as banks and pension funds to pledge not to invest money in the development or manufacture of AI for military use; and
- Activists, academics, affected communities, and other concerned about privacy rights, digital dehumanisation, environmental and climate justice, gender-based violence, and other issues to collaborate and strategise to oppose the development and use of AI in the military and other violent domains.

D. Scientific Community

AI, Automated Systems, and Resort-to-Force Decision Making Research Project, The Australian National University

[11 April 2025]

Introduction

This executive summary highlights policy recommendations outlined in ***AI, Automated Systems, and Resort-to-Force Decision Making – Policy Recommendations: Submission to the UN Secretary General Pertaining to A/RES/79/239*** **(11 April 2025)**, available on the UNODA website. For a complete account of the underlying research and associated research papers, please refer to the full submission.

Underlying Research Project

This research has arisen from a **two-and-half-year research project (2022-2025),** entitled ***Anticipating the Future of War: AI, Automated Systems, and Resort-to-Force Decision Making***, led by Professor Toni Erskine (Australian National University) and funded by the Australian Government through a grant by the Department of Defence.

Its focus is **distinctive and critical**. While the attention of academics and policy makers has been overwhelmingly directed towards the use of AI-enabled systems in the *conduct of war* – including, prominently, on the emerging reality of 'lethal autonomous weapons systems' ('LAWS'), this project has addressed the **relatively neglected prospect of employing AI-enabled tools at various stages and levels of deliberation over the *resort to war.*** In other words, 'it takes us from AI on the battlefield to **AI in the war-room'**.[1]

This research project has brought together **leading scholars and practitioners** working on different aspects of international politics and security, strategic and defence studies, and artificial intelligence (AI) to contribute to a multi-disciplinary study and set of **policy recommendations on the risks and opportunities of introducing AI, machine learning (ML), and automated systems** into state-level decision making on the **initiation of war**. Our interventions are made from the perspectives of political science, international relations, law, computer science, philosophy, sociology, psychology, engineering, and mathematics.

Project participants presented and discussed their research at two workshops (June 2023 and July 2024) at the Australian National University (ANU), convened by Professor Toni Erskine and Professor Steven E. Miller (Harvard). Participants also received feedback on their initial research-based policy recommendations from senior Australian Government delegates from the federal civil service as part of a one-day policy roundtable (July 2024) at the ANU.

'Four Complications'

For all the potential **benefits** of AI-driven systems – which are able to analyse vast quantities of data, make recommendations and predictions by uncovering patterns in data that human decision makers cannot perceive, and respond to potential attacks with a speed and efficiency that we could not hope to match – challenges abound. Through this project, we have sought to address **four thematic**

[1] T. Erskine and S. E. Miller, 'AI and the Decision to Go to War: Future Risks and Opportunities', *Australian Journal of International Affairs*, Vol. 78: 2 (2024), pp. 135–147 (p. 138).

'complications' that we propose will accompany the gradual infiltration of AI-enabled systems in **decisions to wage war:**[2]

- **Complication 1** relates to the displacement of human judgement in AI-driven resort-to-force decision making and possible implications for deterrence theory and the unintended escalation of conflict.
- **Complication 2** highlights detrimental consequences of automation bias, or the tendency to accept without question computer-generated outputs – a tendency that can make human decision makers less likely to use (and maintain) their own expertise and judgement.
- **Complication 3** confronts algorithmic opacity and its potential effects on the democratic and international legitimacy of resort-to-force decisions.
- **Complication 4** addresses the likelihood of AI-enabled systems impacting organisational structures and chains of command, whether degrading or enhancing strategic and operational decision-making processes.

Contributors to this project have explored these proposed complications in the context of either **automated self-defence** or the use of **AI-driven decision-support systems (DDS)** that would inform human resort-to-force deliberations. We have identified risks and opportunities of using AI-enabled systems in these contexts and make recommendations on how risks can be mitigated and opportunities promoted.

Complication 1: Displacement of human judgement

AI in Nuclear Crisis Decision Making

One key area of research undertaken in response to this complication is the nuanced interplay between AI and human decision making in the high-stakes context of **nuclear crisis management**. **Risks (including the increased fragility of nuclear deterrence relationships, crisis signalling becoming more complex**, and **unintended escalation)** have been explored in two broad areas: i) automation in military deployments, or taking the human 'out of the loop' in the decision to use nuclear or strategic non-nuclear weapons (SNNW); and, ii) the integration of AI into human decision-making (particularly in early warning threat assessments). Although much of this research has focused on risks, **novel benefits** of introducing AI-driven decision-support systems (DSS) into human-led nuclear crisis management have also been proposed.

Policy Recommendations:

- **Always incorporate human-in-the-loop safeguards**: Ensure AI systems in nuclear command and control are always overseen by human operators and that human decision-making remains central to determining when and how nuclear-weapon states resort to the use of their arsenals.
- **Promote a holistic approach to AI-safety**: AI safety should account for both technical and socio-technical dimensions. Assess safety challenges in AI-enabled DSS comprehensively, including issues of security, trust, and liability.
- **Broaden the scope of risk assessments**: Apply risk assessments relating to the deployment of AI and ML not only to obvious areas such as nuclear launch orders, but also to less obvious areas such as early warning intelligence

[2] For an account of these 'four complications', see T. Erskine and S. E. Miller, 'AI and the Decision to Go to War: Future Risks and Opportunities', *Australian Journal of International Affairs,* Vol. 78: 2 (2024), pp. 135–147 (pp. 139–40).

assessments (including by non-nuclear allies) and SNNW capabilities (including by non-nuclear allies).

- **Restrict the use of AI-assisted warning data**: The key to balancing the benefits of incorporating AI into early warning against the risks is limiting what AI-assisted warning data is used for. In AI research, prioritise tasks such as calculating effective evasive manoeuvres in the event of an attack and using pattern recognition and anomaly detection to improve arms control verification.
- **Pursue** informal arms control and confidence-building: Advance informal measures such as regular dialogue, red-line agreements, and information-sharing mechanisms. Expand unilateral initiatives like moratoriums where feasible.
- **Explore AI's potential to promote empathy and enhance decision making**: Decision makers must exercise 'security dilemma sensibility' (SDS) in times of crisis. Decision makers and diplomats exercise SDS when they are open to the possibility that the other side is behaving the way they are because they are fearful and insecure, and crucially, recognize the role that their own actions may have played in this. Explore ways that the balanced integration of AI and human judgement could enhance SDS during nuclear crises by promoting empathy and trust.

AI Mistakes in the Resort to Force

Another area addressed in relation to this complication is **state responsibility** when **errors** occur in AI-driven or autonomous systems involved in resort-to-force decisions. Such errors may arise from poor system training, data poisoning by adversaries, or two AI-driven systems interacting in unintended ways. It is essential to develop legal standards and practices that reduce the risk of unintended conflict resulting from such failures.

Policy Recommendations:

- **Adopt robust security and cyber hygiene**: States should adopt robust protections against AI data poisoning and cyber attacks to meet jus ad bellum standards of good faith and reasonable conduct.
- **Clarify legal guidelines on delegating the use of force to autonomous systems**: Senior leadership within states should set clear domestic legal standards regarding when and how autonomous systems may be authorised to use force.
- **Commit to transparency in after-action reviews**: States should commit to being transparent and deliberate about after-action reviews of any AI errors that occur in the field, potentially drawing on civilian casualty review processes as a model.

Complication 2: Automation bias

Our research in response to the second complication focuses on the relationship between human actors and **AI-driven DSS** in resort-to-force decision making. It includes a detailed survey-based study of **military trust in AI** during strategic-level deliberations and a robust account of the importance of ensuring that there are human **'experts-in-the-loop'** when AI-driven systems contribute to decisions on war initiation. This body of work also addresses the **benefits** of employing DSS to **enhance our cognitive capacities** in strategic decision making and, conversely, uncovers the potential **dangers** of such reliance if these systems **dull our sensitivity**

to the tragic qualities of war or contribute to the **erosion of restraint** by creating the illusion that they replace us as responsible actors.

Policy Recommendations:

- **Consider the multidimensionality of trust**: Recognize that soldiers' trust in AI is not a forgone conclusion. Rather, it is complex and multidimensional, and further complicated by biases, uncertainty, and lack of education.
- **Interrogate norm compliance**: In terms of governance, explain how policies on increasingly autonomous capabilities coincide or diverge from international norms and laws informing their use.
- **Embed experts in decision structures**: Enshrine an 'expert-in-the-loop' organisational structure – i.e., high-level experts as core decision makers.
- **Prohibit automation**: Prohibit automation of resort-to-force decisions.
- **Increase AI literacy of domain experts**: Provide and require basic technical training for high-level domain experts so they understand the logics of AI and can thus incorporate AI decision inputs from an informed position.
- **Provide on-going, substantive training for domain experts**: Sustain substantive training for, and assessment of, high-level experts to bolster and ensure substantive competencies.
- **Regulate non-autonomous AI**: While autonomous AI agents, e.g., lethal autonomous weapons systems (LAWS), need regulation, so do non-autonomous AI systems, which leave humans vulnerable to new forms of influence, moral and cognitive atrophy, and undermined responsibility.
- **Design AI-driven DSS to promote more accurate perceptions of their capacities**: Ensure AI-driven DSS are not easily mistaken for responsible agents in themselves by avoiding anthropomorphism, building in warnings about system limitations, and incorporating features that emphasise human agency and accountability.

Complication 3: Algorithmic opacity

Our research in response to the third complication addresses **how the lack of transparency of AI-driven decision making can threaten the legitimacy** of AI-informed decisions on the resort to force. This body of work includes original research on **large language models (LLMs)** and their potential to exacerbate existing **pathologies in intelligence analyses**. It also examines the role that the **'architecture of AI'** and its hidden vulnerabilities play in deliberations surrounding the resort to force. Moreover, research within this pillar conceives of military decision-making institutions as **'complex adaptive systems'** – a conceptual framework that yields a range of insights, including that human-machine teams possess a form of **'cognitive diversity' that could be leveraged for more efficient decision-making,** but also **exploited to poison information flows**, and that technical explanations for algorithmic opacity will not solve accountability concerns.

Policy recommendations:

- **Develop policy to limit epistemic pathologies of LLMs**: States should clearly determine defence and intelligence policy towards either a) procurement of LLMs, b) state development of LLMs, or c) a combination of both. They should use this guidance to develop policy which seeks to limit the epistemic pathologies of LLMs in autonomous decision-making.

- **Commit to sector-wide procurement guidelines and oversight of generative AI tools** used in decision-making chains.
- **Commit to regulating data markets** and access to those markets through alliance relationships.
- **Promote understanding of the tech ecosystem and its fragilities**: Increase understanding of the inherent interdependencies and vulnerabilities in the tech ecosystem, including by creating technology literacy training programs designed specifically for politicians and policy, intelligence, and military leaders.
- **Invest in research** to develop a comprehensive picture of the architecture – physical and digital – that underpins AI, including critical dependencies and vulnerabilities and how access and power are distributed.
- **Invest in research on social media** and its impact on functions of government, including its potential to disrupt democracies, facilitate foreign interference, and influence decision making on the use of force.
- **Recognize AI's current influence**: Significantly increase awareness of government reliance on the architecture of AI, especially for critical government functions, including resort-to-force decision making.
- **Invest in research and development** to maximize the benefits of human-machine cognitive diversity.
- **Implement responsible AI governance programs** that carefully balance accountability with operational efficiency.
- **Perform regular red-team exercises** to ensure that the integration of AI in decision-making institutions does not induce systemic blind spots and vulnerabilities in military decision-making.

Complication 4: Impact on organisational structures

Our research regarding the fourth complication explores both the **beneficial and damaging effects that AI-driven systems can have on institutional structures** in the context of resort-to-force decision making. Studies focus on how AI-driven **DSS can improve 'adaptive culture'** within military organisations, thereby improving wartime decisions, and how the urgent need to **upgrade AI literacy and educate human analysts** should lead us to **reform institutional structures and cultures**. The novel notion of **'proxy responsibility'** is proposed as an institutional response to ensure that responsibility can be meaningfully assigned to humans for resort-to-force decisions that are informed by AI systems. Moreover, original research highlights the significance of the **neglected category of AI 'integrators'** – sandwiched between the 'developers' and 'users' of AI within organisational structures – when it comes to strategic military applications of AI.

Policy recommendations

- **Set (and evolve) measures of effectiveness**. If AI-enabled adaptive capacity is to work effectively, measures of military effectiveness must guide which direction adaptation might take. Establish such measures at the tactical (battlefield) and strategic (war-room) levels to guide development and implementation of AI-enabled adaptation.
- **Know where adaptation relevant data is found, stored and shared**. An enhanced adaptive stance in military institutions must have enhanced data

awareness as a foundation. Data awareness and management must become one of the basic disciplines taught to military personnel.

- **Scale AI support from individual to institution**. There is unlikely to be a one-size-fits-all algorithm or process that can enhance learning and adaptation at every level of military endeavours. Create a virtual ‘arms room’ of adaptation support algorithms as part of an institution-wide approach to adaptation.
- **Routinely question AI-enabled outputs**: Build mindsets, protocols, institutional cultures, and inter-agency structures in ‘normal’ pre-crisis times to routinely question AI-enabled output from human-machine teams.
- **Institute an advisory body**: In order to support the notion of ‘proxy responsibility’ as an institutional response to ‘responsibility gaps’ when decisions on war initiation are informed by AI-enabled systems, establish and/or strengthen state-level ‘AI departments’. These departments would integrate technical, political, and ethical competence and expertise and advise on resort-to-force decision-making processes.
- **Support research on AI integration**: Fund research on the integration of AI in strategic decision-making.
- **Provide standards**: Outline minimum standards for the responsibilities of AI developers and integrators.
- **Facilitate inter-group discussions** between developers, integrators and users during development, integration, and longer-term maintenance processes.
- **Create accountability guidelines**: Provide well-defined guidelines and rules indicating who is accountable if something goes ‘wrong’.

Queen Mary University of London, T.M.C Asser Institute, University of Southern Denmark, University of Utrecht

[11 April 2025]

Views of members of the scientific community and civil society; specifically, we are a group of academics with expertise in ethical, legal and political dimensions of military Artificial Intelligence and herewith put forward our shared views pursuant to resolution 79/239 “Artificial intelligence in the military domain and its implications for international peace and security” adopted by the General Assembly on 24 December 2024, in accordance with the request of the UN Secretary-General contained in Note Verbale ODA/2025-00029/AIMD.

Introduction:

The rapid advancement and integration of AI technologies into targeting operations have sparked ongoing debates surrounding their ethical, legal, and operational implications. Over the past decade, the discourse on AI in warfare has largely centered on autonomous weapon systems (AWS),[1] driven in part by the initiation of discussions in 2013 and the formalization of a regulatory process under the UN Convention on Certain Conventional Weapons (CCW) and the Group of Governmental Experts on Lethal Autonomous Weapons Systems (GGE LAWS),

[1] The latest definition of AWS from the CCW GGE LAWS Rolling Text (26 November 2024): “A lethal autonomous weapon system can be characterized as an integrated combination of one or more weapons and technological components that enable the system to identify and/or select, and engage a target, without intervention by a human user in the execution of these tasks.” On file with authors.

which exclusively focuses on lethal AWS.[2] However, the increasing integration of AI-based decision-support systems (AI-DSS) into targeting practices[3] introduces new layers of complexity that demand closer attention from a broad range of stakeholders. This submission responds to that need, structured around three key components: (1) a brief overview of how AI-DSS are currently used in targeting decisions; (2) an analysis of key concerns, including how these systems shape the potential exercise of human judgement and control and underline fundamental gaps in global governance; and (3) a concluding set of recommendations.

1. Overview of AI-DSS and the joint targeting cycle

Defined as "the process of selecting and prioritizing targets and matching the appropriate response to them, considering operational requirements and capabilities,"[4] targeting is a core military function at the very heart of warfare. While the potential range of use cases for AI-DSS in military decision-making is broad, in targeting, AI-DSS can be understood to serve as **tools** that use AI techniques to collect and analyze data, provide information about the operational environment as well as actionable recommendations, with the aim of aiding military decision makers in evaluating factors relevant to legal compliance such as taking precautions and ensuring proportionality in attacks.

More specifically, AI-DSS are increasingly integrated across multiple phases of the joint targeting cycle (JTC), including within target development and prioritization, capabilities analysis, and mission execution. The JTC is a reflective example of a structured process used by military forces to identify, evaluate, and engage targets while ensuring compliance with operational, legal, and ethical standards,[5] generally consisting of six (non-linear) phases:

1. **End-State and Commander's Objectives**: Defining strategic military goals and desired outcomes.

---

[2] For a brief overview of some of the latest developments of the GGE LAWS see Jeroen van den Boogaard, *Warning! Obstacles Ahead! The Regulation of Autonomous Weapons Systems in the GGE LAWS*, *Opinio* Juris, 4 March 2024 found at: https://opiniojuris.org/2024/03/04/warning-obstacles-ahead-the-regulation-of-autonomous-weapons-systems-in-the-gge-laws/.

[3] There have been several reported uses of AI-DSS by Israel in Gaza and potentially in Lebanon, by both Ukraine and Russia in the ongoing conflict, and by the United States in its actions against Houthi rebels in the Red Sea and in Yemen, to name a few. For a comprehensive overview of literature in this space, see e.g., Anna Nadibaidze, Ingvild Bode, and Qiaochu Zhang, *AI in Military Decision Support Systems, A Review of Developments and Debates,* Centre for War Studies, University of Southern Denmark, November 2024. Found here: https://www.autonorms.eu/ai-in-military-decision-support-systems-a-review-of-developments-and-debates/.

[4] United States Department of Defense, *Dictionary of Military and Associated Terms*, March 2017, found at: https://www.tradoc.army.mil/wp-content/uploads/2020/10/AD1029823-DOD-Dictionary-of-Military-and-Associated-Terms-2017.pdf.

[5] Michael Schmitt et al, *Joint and Combined Targeting: Structure and Process*, Chapter 13 in Jens David Ohlin (ed) *Weighing Lives in War* (Oxford, 2017). See also, Jessica Dorsey and Marta Bo, *AI-Enabled Decision-Support Systems in the Joint Targeting Cycle: Legal Challenges, Risks, and the Human(e) Dimension*, forthcoming 2025, *International Law Studies.* "Targeting generally involves four key steps: (1) objectives and guidance, (2) planning, (3) execution, and (4) assessment. Encapsulating these four key steps, the United States and NATO outline their targeting processes through similar six-phase cycles [addressed in this submission]. As the reader can discern, different states employ different doctrines for targeting. What is important … is not necessarily the specific labels for various steps followed by any given state, but rather how and when compliance with the principle[s of IHL are] incorporated into the targeting process."

2. **Target Development and Prioritization**: Identifying, verifying/validating, and prioritizing targets based on intelligence and mission goals.

3. **Capabilities Analysis**: Assessing the available strike options and their effectiveness.

4. **Force Assignment**: Allocating specific military assets (e.g., airstrikes, artillery, cyber operations) to engage the target.

5. **Mission Execution**: Carrying out the targeting operation while ensuring compliance with relevant laws and the rules of engagement.

6. **Assessment**: Evaluating the effectiveness of the operation and adjusting for future operations, if necessary.

Within this framework, AI DSS are assumed to serve primarily as informational and analytical tools which support human decision-making rather than supplant it. However, this assumption and framing obscures how AI-DSS influence human cognitive processes within the JTC. This impact on human decision-making is often underestimated and remains insufficiently examined, leaving critical discussions about the role of AI-DSS largely absent from current policy debates.

2. Analysis of Key Concerns

(a) (Meaningful) Human Judgement and Control

AI-DSS are often portrayed as enhancing human decision-making and the quality of decisions therein. The perception of AI-DSS as mere subsidiary tools has led to a narrative that the integration of AI-DSS poses fewer challenges than AWS, given that these systems do not directly "engage" targets (i.e., they do not have an inherent capability to directly carry out the use of force) and are tools that assist human commanders. The outputs are ostensibly ultimately reviewed through (several layers of) human oversight, such as processes of verifying and validating targets using additional intelligence sources.[6] As a result, errors or inaccuracies in AI-DSS outputs are often seen as non-critical, based on the assumption that robust human oversight and appropriate control will compensate for them. However, closer examination reveals that this control is frequently superficial, offering only the appearance of, rather than actual meaningful, or context-appropriate, human judgement and control.

This is because AI-DSS structure and condition the quality of human control and oversight and limit the ways control and oversight can be exercised. The use of AI-DSS creates a shared decision-making space between human military personnel and AI technologies. States appear to have recognized and focused on many of the advantages of this shared decision-making space for military personnel, i.e., how the use of AI-DSS advances human decision-making through offering data-driven insights. But using AI-DSS also delimits the capacity to exercise human oversight and control because of the technologies' complexity and the increased speed (and therefore scale) it can bring to decision-making processes. Rather than supporting human oversight, using AI-DSS may risk humans becoming little more than reactive cogs in socio-technical systems.[7] Moreover, this configuration risks amplifying adverse human biases, such as automation bias, anchoring bias, or cognitive action bias, to the detriment of exercising qualitatively high levels of human control.[8]

---

[6] Alexander Blanchard and Laura Bruun, *Automating Military Targeting: A Comparison Between Autonomous Weapon Systems and AI-Enabled Decision Support Systems*, Stockholm International Peace Research Institution (SIPRI) forthcoming 2025 (draft on file with authors).

[7] Ingvild Bode, *Human-Machine Interaction and Human Agency in the Military Domain*, Policy Brief No. 193 (Waterloo, ON: Centre for International Governance Innovation, 2025).

[8] Dorsey and Bo, supra n. 5.

Considering AI-DSS as a distinct form of technology therefore reveals significant challenges associated with military AI and human oversight, challenges that extend beyond those that arise when simply integrating the technology in weapon systems.

Recent conflicts have shown the risks associated with AI-DSS being employed in critical functions, such as target selection and even nomination, and their conditioning and constraining of human involvement, affecting the fulfilment of core legal obligations embedded within the JTC. The use of AI-DSS raises fundamental concerns about whether human decision makers can retain adequate cognitive autonomy over the JTC process or whether humans will become overly reliant on algorithmic outputs for critical judgements in the context of armed conflict.[9] Consequently, there are significant legal concerns regarding the effects of such systems on decision-making processes and use of force decisions and ability for users to comply with IHL obligations, especially with respect to the obligation to take all feasible precautions to minimize civilian harm to the greatest extent possible in attack and comply with the principles of distinction and proportionality.[10]

Importantly, these concerns are not new. There is extensive debate around how to preserve meaningful human judgment and human agency when conducting IHL-evaluative legal assessments, in the context of AWS. These discussions – which include expert analysis on accountability, human-machine interaction, automation bias, and the effect of AI systems on legal and ethical reasoning[11] – provide valuable lessons that must inform discussions around military AI and specifically the use of AI-DSS in military contexts.

(b) AI-DSS: Understudied, Under-Addressed and Unregulated

Framing AI-DSS as mere tools, has led to an underestimation and lack of analysis on the way their use affects the cognitive decision-making process within the JTC. The relative lack of attention paid to AI-DSS so far can partly be attributed to the fact that such systems are seen to be used with a human *in* or *on* the loop framework, with their outputs ostensibly reviewed by one or more individuals during the targeting process. As a result, current understandings of AI-DSS use appear to align with widely supported principles of human control and oversight. However, this gap in the debate

---

[9] *Ibid*; see also Anna Nadibaidze, Ingvild Bode, and Qiaochu Zhang, *AI in Military Decision Support Systems, A Review of Developments and Debates,* Centre for War Studies, University of Southern Denmark, November 2024. Found at: https://www.autonorms.eu/ai-in-military-decision-support-systems-a-review-of-developments-and-debates/.

[10] Article 57 of the First Additional Protocol to the Geneva Conventions. See also Dorsey, Bo supra n. 5 (on AI-DSS and their effects on the principle of precautions); Jessica Dorsey, *Proportionality under Pressure: The Effects of AI-Enabled Decision Support Systems, the Reasonable Commander Standard and Human(e) Judgment in Targeting*, forthcoming *International Review of the Red Cross* (2025) (on AI-DSS and their effects in the context of IHL proportionality assessments).

[11] Marta Bo, *Autonomous Weapons and the Responsibility Gap in light of the Mens Rea of the War Crime of Attacking Civilians in the ICC Statute*, 19 Journal of International Criminal Justice 2021; Bo, M., Bruun, L. and Boulanin, V., *Retaining Human Responsibility in the Development and Use of Autonomous Weapon Systems: On Accountability for Violations of International Humanitarian Law Involving AWS* (SIPRI: Stockholm, Oct. 2022), p. 41; Boulanin, V., Bruun, L. and Goussac, N., *Autonomous Weapon Systems and International Humanitarian Law: Identifying Limits and the Required Type and Degree of Human-Machine Interaction* (SIPRI: Stockholm, June 2021), p. 54; and Bruun, L., Bo, M. and Goussac, N., *Compliance with International Humanitarian Law in the Development and Use of Autonomous Weapon Systems: What Does IHL Permit, Prohibit and Require?* (SIPRI: Stockholm, Mar. 2023), p. 24. Elke Schwarz, "The (im)possibility of meaningful human control for lethal autonomous weapons systems," *Humanitarian Law and Policy*, 29 August 2018, found at: https://blogs.icrc.org/law-and-policy/2018/08/29/im-possibility-meaningful-human-control-lethal-autonomous-weapon-systems/.

is also caused by a lack of transparency around how specific AI-DSS function, and a consistent failure to comprehensively examine how they are being used in practice.

Additionally, the persistent focus on AWS at the expense of AI-DSS obscures the growing reliance on AI in shaping operational and strategic outcomes. Unlike AWS, which have been debated in the framework of the CCW for the past decade, AI-DSS lack a comparable institutional platform. Attention to AI-DSS remains scattered across various initiatives but these efforts have yet to provide the dedicated regulatory focus or coordination needed.

3. **Recommendations:**

i. **Reassert** the central role of human cognitive and legal reasoning in military operations by implementing safeguards that ensure key legal assessments remain grounded in human(e) judgment. Leverage existing insights from debates on AWS and research on human-machine teaming and human-computer interaction to inform discussions on AI-DSS.

ii. **Recognize** and address the incremental effects of AI-DSS design and use on human cognitive reasoning and critical deliberation. Promote awareness and attentiveness as a crucial part of reasserting and strengthening the exercise of human agency in targeting decision-making.

iii. **Reinforce** calls for greater attention to the implications of AI-DSS in armed conflict. Utilize platforms such as the UN General Assembly's First Committee on Disarmament and International Security and the Global Commission on the Responsible Use of AI in the Military Domain to foster inclusive and complementary discussions on the associated risks and systemic changes AI-DSS introduce.

United Nations Institute for Disarmament Research

[11 April 2025]

Artificial intelligence (AI) is rapidly transforming the military domain and profoundly influencing international peace and security. Initiatives such as the summits on Responsible AI in the Military Domain (REAIM) and the Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, while not being universal processes, have significantly elevated international attention on the military applications of AI. In particular, they have moved the debate beyond lethal autonomous weapon systems (LAWS) and have successfully highlighted the multifaceted impacts of AI, fostering broader international policy engagement. Building on the political momentum generated by these initiatives, resolution 79/239 adopted by the United Nations General Assembly in December 2024 represented a significant milestone as the first UN resolution on AI in the military context and has offered Member States, international and regional organizations and the multi-stakeholder community the opportunity to share their views on opportunities and risks.

For many years, the United Nations Institute for Disarmament Research (UNIDIR) has played an important role in shaping and informing discussions on the broader impact of AI in the military domain, both within and beyond applications of this technology in weapon systems. It has undertaken research, facilitated multilateral dialogues, and offered policy insights that underline AI's transformative potential for international peace and security. This policy note draws from all the work conducted to summarise opportunities and risks and to offer a potential roadmap for future policy action.

The international community can now shape how AI is used in the military domain, putting principles of responsible AI at the core. A central challenge is the

complexity of defining the “military domain”. States and regions interpret the scope of this domain differently based on their unique security landscapes, realities and operational practices. For some countries, military roles extend to internal security tasks such as policing, border control, combating organized crime, protection of critical infrastructure or humanitarian relief in response to natural disasters. Others maintain a stricter definition, limiting military functions to battlefield engagements. These variations, rather than serving as barriers, offer important context for multilateral discussions. International governance frameworks should remain flexible and inclusive, acknowledging and adapting to diverse national and regional security perspectives.

In the many operational contexts within the military domain, AI acts as a force multiplier across several military tasks, including command and control (C2), information management and intelligence, advanced autonomy, logistics, training and simulation, and organizational and support functions. In C2, AI enhances the speed and quality of decision-making, thereby helping commanders rapidly analyse battlefield scenarios. It has the potential to improve adherence to international humanitarian law (IHL), for example by integrating detailed proportionality and other legal assessments. AI-driven intelligence tools analyse large volumes of data at speed, and so improve situational awareness and threat detection. In logistics, AI optimizes supply chains and predictive maintenance, enhancing operational readiness and improving the sustainability of military operations over time. AI further supports advanced autonomy in drones, cybersecurity, and operations in the information domain. Training and simulation benefit from AI by creating personalized, realistic synthetic environments and scenarios. In short, if developed, deployed and used responsibly, AI could increase operational effectiveness while offering new ways to mitigate risks and reduce harm.

However, integrating AI in military contexts also presents significant risks and challenges – technological, security, legal, policy and ethical.

Technologically, military AI systems face issues related to the quality, availability and inherent biases of data. These may lead to unpredictable and potentially harmful outcomes, including violations of international law. The “black box” nature of AI systems, often coupled with their adaptiveness and highly context-dependent nature, complicates trustworthiness assessments and may, at times, challenge the conduct of effective investigations into alleged violations of IHL. Cybersecurity vulnerabilities also expose AI systems to adversarial attacks, requiring stringent security measures.

Security challenges include risks of miscalculation and unintended escalation, particularly through AI-enabled rapid decision-making processes and AI-enabled autonomy, which may result in escalatory responses. The potential for an AI arms race exacerbates international and regional tensions, possibly leading to destabilizing outcomes similar to historical arms competitions. The proliferation of AI technologies to non-state actors further complicates threat landscapes and necessitates robust life cycle management of military AI systems. Additionally, AI-generated disinformation threatens societal stability by undermining trust in information and can have a direct impact on military operations.

Legal challenges revolve around ensuring compliance with international law, particularly IHL and international human rights law. Key debates focus on, among other things, accountability and both state and individual responsibility for AI-driven actions, especially regarding lethal decisions. States diverge on whether existing legal frameworks are sufficient or if new, specialized regulations are required. Beyond international law, ethical considerations emphasize maintaining human judgment in critical decision-making and preventing societal biases from infiltrating AI systems. The latter requirement calls for greater diversity and inclusivity in AI development. Additionally, bridging gaps between government, academia and the private sector remains challenging yet crucial for effective governance.

Addressing these challenges requires a comprehensive road map with actions at the multilateral, regional and national levels.

Multilaterally, establishing a United Nations-led comprehensive platform that enables a regular institutional dialogue to address military AI's broader implications on international peace and security is key as it would provide an institutional framework to advance policy discussions. This platform could build on the existing internationally developed AI principles and frameworks, such as UNESCO's recommendations or the commitments made in the Global Digital Compact (e.g. safe, secure and trustworthy AI) and further refine them for application in the military domain. These principles could be further developed into voluntary norms of responsible behaviour in the development, deployment and use of AI in the military domain and provide a solid foundation for future multilateral instruments. In addition, such platform could be leveraged to develop practical confidence-building measures (CBMs), lead inclusive multi-stakeholder engagement, and deliver global capacity-building programmes that enhance global security via transparency, cooperation and predictability.

Regionally, existing organizational frameworks can be used to tailor CBMs and guidelines that reflect local security contexts. Cross-regional dialogues would facilitate mutual learning, prevent information silos, and include diverse perspective which would encourage globally coherent responses.

Nationally, states should develop comprehensive AI strategies that detail vision, priorities and governance frameworks, ensuring compliance with international norms and ethical standards. Robust governance structures (e.g., dedicated AI steering committees and ethics boards), alongside iterative legal reviews, would enhance accountability and safety. Transparent communication and clearly defined accountability protocols would further support responsible AI implementation. High standards of data governance, life cycle management approaches, rigorous training programmes and updated military operational guidelines complete these proposed national measures, ensuring the responsible integration of AI in the military domain.

Table A below provides an overview of the proposed roadmap for policy action.

Table A: A roadmap for future policy action

Level	*Action*	*Rationale*
Multilateral	**Establish a multilateral process under United Nations auspices to provide a comprehensive platform for discussion on military applications of AI and their impact on international peace and security.** This process could be leveraged to: a. Develop a set of overarching, core principles of responsible AI in the military domain to help align national efforts and reduce risk. b. In the future, further develop these core principles into international voluntary norms or guidelines for responsible state behaviour in the	Collectively, these multilateral actions aim to foster cooperation, set common rules and share knowledge on military AI at the international level with a view to increasing predictability. They aim to shape the global landscape so that all states move towards safer and more transparent integration of AI in the military domain, thereby reducing the risks. While clustered under a single umbrella recommendation, each of the actions above could be implemented on its own, although their mutually reinforcing nature would amplify

Level	*Action*	*Rationale*
	development, deployment and use of AI in the military domain. These guidelines could take the form of a code of conduct or a political declaration supplemented by more technical instruments as required (e.g., on AI assurances, and robust protocols for testing and evaluation).	the impact achieved if they are implemented in combination.
	c. Develop confidence-building measures (CBMs) for military AI. States could agree on and implement practical CBMs to increase transparency and trust regarding AI in the military domain.	
	d. Promote multi-stakeholder engagement in support of multilateral policy action.	
	e. Develop and implement a coherent capacity-building programme.	
Regional	**Leverage regional and subregional organizations and dialogues to discuss the issue of AI in the military domain.** Regional and sub-regional organizations could:	Regional and subregional approaches allow tailoring to specific security realities and threat perceptions, which could lead to concrete results that are more aligned with specific needs.
	a. Develop region-specific CBMs, norms or guidelines that reflect local contexts.	In addition, regional and subregional approaches could be leveraged to inform and shape global dialogues and strengthen context-specific capacity-building.
	b. Set up networks for information-sharing on AI-related best practices suited to their security landscape.	Cross-regional dialogue can be a useful tool to enable mutual learning and avoid echo chambers.
	c. Develop joint AI-development projects, aligning operational, legal and technical requirements.	
	Initiate cross-regional dialogues Initiate cross-regional dialogues on AI, where two or more regional groups exchange lessons and possibly align their approaches.	
National	**Implement a comprehensive approach to AI governance in**	A national strategy clarifies roles and responsibilities, and provides

Level	*Action*	*Rationale*
	the military domain to include the following actions: Develop a comprehensive national strategy or policy on AI in security and defence. b. Establish robust governance structures and review processes. c. Implement transparency and accountability measures d. Implement robust data practices and governance frameworks for all military AI applications. e. Manage AI capabilities throughout their entire life cycle – from design and development, through testing and deployment, to updates and decommissioning – with continuous risk assessments and mitigation at each stage. f. Invest in human capital and training by developing extensive training programmes for military personnel on AI and cultivating a new generation of AI-literate officers and specialists. This includes not only technical training but also training on the ethical and legal aspects of AI use in operations. g. Review military operational guidelines to strengthen AI governance in military contexts, including military documentation (e.g. doctrines, standard operating procedures and others), and rules of engagement.	a clear direction for the development, acquisition, integration and use of AI in the military domain. Dedicated structures provide focus and accountability. They create effective checkpoints that AI projects must pass and comply with consistently (e.g., ethical approval, legal clearance, safety certification), reducing chances of unsafe or unlawful deployment. Transparency builds public trust and international confidence that a state is using AI responsibly. Accountability ensures that the presence of AI does not create a vacuum of responsibility – maintaining the ethical and legal norm that humans are accountable for military actions. By prioritizing robust data governance and the provision of the necessary infrastructure to enable it, militaries can improve the performance and trustworthiness of their AI systems and reduce error rates. A life cycle view ensures that safety and compliance are ongoing commitments reducing chances of failure in the field and ensuring that accountability is maintained throughout the system’s use. Human expertise and judgment remain critical. Training reduces misuse and enables more effective human-machine teaming. Existing military governance tools and instruments can be used to strengthen the governance of AI in the military domain at a more practical, tactical level, thereby offering an impactful complement to the highest levels of governance and

Level	*Action*	*Rationale*
		the associated obligations emanating from international, regional and national laws and regulations.

E. Industry

Microsoft

[24 May 2024]

Microsoft welcomes the opportunity provided by the United Nations General Assembly resolution A/RES/79/239 on "Artificial Intelligence in the Military Domain and its Implications for International Peace and Security", and UNODA's invitation to share perspectives on the opportunities and challenges posed to international peace and security by the application of artificial intelligence (AI) in the military domain, with specific focus on areas other than lethal autonomous weapons systems.

Our perspectives reflect Microsoft's deep commitment to our Responsible AI Principles and our Secure Future Initiative, emphasizing cybersecurity, safeguarding international norms, and promoting trust in technology, and our active participation in multi-stakeholder initiatives including the UNIDIR-led Roundtable for AI, Security, and Ethics (RAISE).

I. Opportunities

Microsoft recognizes substantial opportunities in responsibly applied AI within the military domain, particularly:

- *Enhancing cybersecurity and defense capabilities:* AI significantly strengthens cybersecurity defenses by automating threat detection, enabling faster and more accurate responses to cyber threats. Technologies such as Microsoft Security Copilot illustrate the transformative potential of AI in defense, empowering cybersecurity professionals to identify and mitigate risks efficiently. Initiatives like Microsoft's Zero Day Quest and collaboration with MITRE ATT&CK demonstrate proactive industry efforts to enhance global cybersecurity preparedness and resilience.
- *Broad spectrum of military applications:* Beyond cybersecurity, responsibly designed AI can significantly enhance efficiency and effectiveness across logistics, command and control systems, intelligence processing, military training, peacekeeping, humanitarian assistance, and disaster relief operations. Diverse applications underscore AI's transformative potential beyond combat scenarios alone.
- *Improving compliance with international humanitarian law:* AI technologies should improve the accuracy and effectiveness of targeting processes, aiding militaries to better adhere to principles of distinction, proportionality, and necessity. AI should significantly enhance protections for civilians and civilian infrastructure, thereby reducing unintended collateral damage in conflict.
- *Capacity building and international cooperation:* The adoption of AI in the military domain presents opportunities for global knowledge-sharing and capacity-building initiatives. International partnerships should support

developing nations by sharing security capabilities, knowledge, and best practices, thus bridging technological divides and fostering global stability.

II. Challenges

Microsoft equally acknowledges significant challenges and risks associated with AI applications in the military domain:

- *AI-enhanced cyber threats:* AI has escalated cyber threat capabilities, empowering state-sponsored and criminal actors to carry out increasingly sophisticated cyber operations. These AI-driven threats include advanced phishing campaigns, automated exploitation of vulnerabilities, and adaptive malware, significantly increasing global cybersecurity risks.
- *Risks of escalation and miscalculation:* Integrating AI into military decision-making risks unintended escalation and/or miscalculation. Rapid, automated decision-making processes may inadvertently lower conflict thresholds, amplifying risks of destabilization or accidental conflict.
- *Proliferation and uncontrolled diffusion:* Uncontrolled diffusion, especially through open-source models and decentralized development, heightens the risk of malicious use by both state and non-state actors, including terrorist groups and cyber mercenaries. Increasingly accessible dual-use and proprietary AI systems enable actors even with limited resources can gain access to capabilities that previously required significant investment or expertise, posing additional threats to international security and stability.
- *Algorithmic bias and ethical implications:* Algorithmic biases embedded within AI systems pose ethical and humanitarian concerns. Biases related to gender, race, age, or socioeconomic factors in AI datasets can intentionally and unintentionally perpetuate inequality and discrimination, particularly within sensitive military and security applications.
- *Digital divides and inequality:* Without deliberate policy actions, disparities between developed and developing nations in AI capabilities could deepen, increasing geopolitical tensions and socio-economic inequalities, thus undermining long-term global stability.

III. Relevant normative proposals

Microsoft recognizes several existing and emerging normative frameworks relevant to AI governance in the military domain, including:

- UNIDIR’s RAISE initiative, facilitating international multi-stakeholder dialogues and governance proposals.
- The Responsible AI in the Military Domain (REAIM) Summits, emphasizing transparency, accountability, and human oversight at the international level.
- The US Department of Defense Responsible AI Strategy, highlighting responsibility, equitability, traceability, reliability, and governability.
- NATO’s Principles of Responsible Use for AI in Defence, emphasizing reliability, governability, and traceability among member nations.

IV. Microsoft recommendations

To maximize opportunities and mitigate the challenges, Microsoft proposes several key recommendations:

- *Establish clear international norms and standards:* Develop explicit international norms and industry standards governing responsible use and development of military AI. These norms should delineate acceptable and unacceptable behaviors, providing robust frameworks to deter misuse and foster transparency and accountability, supported where appropriate by monitoring or compliance mechanisms. AI governance frameworks should explicitly differentiate operational contexts, such as peacekeeping, humanitarian assistance, crisis management, and conflict scenarios, to appropriately address varied ethical, legal, and humanitarian considerations. To ensure continued relevance, such norms should be periodically reviewed and updated to reflect evolving technological developments and operational realities.

- *Ensure human-centric oversight and accountability:* Adopt policies ensuring meaningful human judgment, oversight, and accountability remain central to military decisions involving AI, particularly regarding the use of force. Clear oversight mechanisms and enforceable accountability structures, including rigorous human control and review processes, are necessary to maintain ethical standards, avoid automation bias, and mitigate unintended consequences.

- *Advance secure and transparent AI development practices:* Promote rigorous technical standards and comprehensive life cycle management protocols covering pre-design, development, testing, deployment, operation, acquisition, and decommissioning. Robust vulnerability management, security audits, and transparent development and deployment processes should be integral components, alongside clear capacity-building measures, ensuring AI systems remain secure, responsible, and resilient throughout their operational life cycle.

- *Enhance responsible data governance practices:* Establish clear international guidelines on responsible data governance specifically tailored to military AI applications. Transparent and accountable data management practices addressing collection, sharing, storage, training, and operational usage are crucial for managing dual-use risks, preventing misuse, and maintaining strict compliance with international legal and ethical frameworks.

- *Address and reduce algorithmic bias:* Prioritize addressing algorithmic bias through rigorous testing, transparent data practices, and inclusive AI development processes. Developers and users should establish clear policies to proactively identify, mitigate, and remediate biases, especially when AI systems are deployed in sensitive military or security contexts.

- *Promote responsible innovation and risk-based regulation:* Support regulatory frameworks that are risk-based, outcome-oriented, and balanced, ensuring they encourage innovation while adequately addressing security and ethical risks associated with AI deployment. Industry should advocate for flexible, adaptive regulations that keep pace with technological change, without imposing overly prescriptive or impractical requirements. Industry-led initiatives, such as voluntary codes of conduct, vulnerability disclosure standards, and collaborative red-teaming exercises, should be actively supported and integrated into broader international normative frameworks.

- *Strengthen international governance and alignment:* Support and actively engage in international initiatives, including REAIM Summits and dialogues at the UN General Assembly and UN Security Council. Robust international governance frameworks, characterized by transparency, clear accountability measures, and trust-building mechanisms, are essential for coherent and inclusive approaches to AI governance. Member States and stakeholders should coordinate closely through these forums to reduce fragmentation and ensure global alignment.

- *Support knowledge-sharing and awareness-raising with the UN system:* Encourage and actively contribute to efforts by the UN Secretariat and relevant UN entities to convene meaningful multi-stakeholder expert dialogues, workshops, and knowledge-sharing on AI in the military domain. Exchanges through voluntary contributions, technical expertise, and collaborative initiatives should aim at enhancing global understanding of AI's implications for international peace and security.
- *Strengthen international cooperation and information sharing:* Encourage robust international cooperation, emphasizing real-time threat intelligence sharing and joint attribution mechanisms. Industry actors should actively participate in collective cybersecurity efforts, enhancing global cybersecurity preparedness and response.
- *Foster multi-stakeholder dialogue and collaboration:* Actively participate in and support forums such as RAISE, involving states, international organizations, academia, civil society, and industry. Such inclusive dialogues are essential for mutual understanding, shaping responsible AI practices, and developing collaborative governance structures.

V. Conclusion

Microsoft is deeply committed to proactive collaboration with Member States, the UN system, industry, and civil society to implement these recommendations swiftly and effectively. Through sustained collective efforts and ongoing engagement in multi-stakeholder initiatives, Microsoft will continue supporting responsible AI governance, innovation, and practices that meaningfully contribute to international peace and security.