GUERRE, PAIX ET ALGORITHMES : LA DISSUASION NUCLÉAIRE FACE AUX DÉFIS DE L'IA

Alain DE NEVE

Airmen and guardians launch an unarmed Minuteman III intercontinental ballistic missile from Vandenberg Space Force Base, Calif., June 4, 2024. The test launch is designed to demonstrate the United States' nuclear deterrent is safe, secure, reliable and effective to deter threats, and reassure allies. @ Air Force Airman 1st Class Olga Houtsma

De opkomst van artificiële intelligentie (AI) werpt nieuwe vragen op over nucleaire afschrikking en strategische stabiliteit. Deze bijdrage onderzoekt hoe AI de besluitvorming binnen nucleaire commandostructuren kan beïnvloeden, met bijzondere aandacht voor de risico's van automatisering, de ethische implicaties en de impact op crisisbeheer. In het huidige geopolitieke klimaat is het cruciaal om het menselijke toezicht te behouden. België en de Europese Unie worden opgeroepen tot een proactieve rol in het reguleren van AI-gebruik in strategische contexten.

Alain DE NEVE est chercheur « Capacités innovantes et défis technologiques dans le domaine de la défense » au sein du Centre d'études de sécurité et défense (CESD) de l'Institut royal supérieur de défense (IRSD).

INTRODUCTION

Bien que souvent perçue comme un phénomène propre au XXIe siècle, l'intelligence artificielle (IA) s'inscrit dans une histoire plus longue, marquée par les transformations successives des technologies militaires et des doctrines de commandement stratégique. Dès les premières spéculations sur la mécanisation du champ de bataille au tournant du XXe siècle, jusqu'aux débats contemporains sur les algorithmes de décision, l'enjeu de la délégation de la décision militaire à la machine n'a cessé d'interroger.

Cette contribution s'attache à analyser les implications concrètes de l'intégration de l'IA dans les systèmes de commandement et de contrôle nucléaire, en replaçant cette évolution dans le sillage des réflexions stratégiques engagées dès la guerre froide. Le propos ne se limite pas à l'étude des potentialités techniques : il s'agit aussi d'évaluer les risques systémiques qu'une telle automatisation fait peser sur la stabilité stratégique, la gestion des crises, et les équilibres de la dissuasion.

Deux lignes de questionnement structurent cette réflexion. La première interroge les conditions dans lesquelles l'IA pourrait être appelée à jouer un rôle opérationnel dans la chaîne décisionnelle nucléaire, ainsi que les limites de cette évolution au regard des principes de contrôle humain. La seconde porte sur les conséquences géopolitiques d'une telle transformation, dans un contexte où les systèmes d'alerte, de ciblage et de riposte deviennent eux-mêmes objets de compétition technologique.

L'exemple du système soviétique Perimetr, tout comme les expérimentations américaines du programme Survivable Adaptive Planning Experiment (SAPE), montrent que ces interrogations ne sont pas neuves. Mais l'accélération actuelle des capacités d'analyse algorithmique, couplée à l'ampleur inédite des données disponibles, redonne une acuité nouvelle à ces débats — en particulier dans ce que l'amiral français Pierre Vandier qualifie de « troisième âge de la dissuasion nucléaire [1]» , marqué par l'irruption des technologies numériques et cognitives au cœur même des postures stratégiques.

Dans ce contexte, la Belgique, bien qu'elle ne dispose pas de l'arme nucléaire, n'est nullement à l'écart de ces enjeux. En tant que membre de l'OTAN impliqué dans la posture de dissuasion élargie et hébergeant des infrastructures critiques relevant du dispositif nucléaire allié, elle est directement concernée par les évolutions doctrinales et technologiques qui affectent les chaînes de commandement intégrées. Par ailleurs, la participation croissante de la Belgique à des projets européens de recherche en défense, y compris dans

[1] Le « troisième âge de la dissuasion nucléaire », selon l'amiral Pierre Vandier, désigne une nouvelle phase stratégique marquée par l'intégration croissante des technologies numériques, de l'IA et de la guerre informationnelle dans les doctrines de dissuasion. Après un premier âge centré sur la parité nucléaire américano-soviétique et un second axé sur la prolifération régionale, ce troisième âge reflète une mutation où les capacités cognitives, la rapidité décisionnelle et la résilience informationnelle deviennent aussi cruciales que les arsenaux eux-mêmes. Pierre Vandier, La dissuasion au troisième âge nucléaire (Paris : Éditions du Rocher, 2024).

le cadre du Fonds européen de la défense (FED), l'expose à des décisions structurantes sur l'usage de l'IA dans les systèmes de commandement et de contrôle. Enfin, à travers son engagement dans les instances multilatérales de gouvernance des technologies émergentes, la Belgique a un rôle à jouer dans la formulation de normes éthiques et juridiques internationales encadrant l'usage de l'IA en matière stratégique.

Ce texte propose donc une analyse des évolutions en cours, tout en mettant en lumière les enjeux éthiques, juridiques et doctrinaux qu'elles soulèvent. Car la perspective d'une dissuasion algorithmisée — ou d'une prise de décision assistée par des systèmes opaques — pourrait profondément redéfinir les fondements même de la sécurité internationale, en remettant en cause l'un des piliers de la guerre nucléaire moderne : la maîtrise humaine sur l'irréversible.

COMMANDEMENT NUCLEAIRE : L'IRRUPTION DE L'IA

Selon John R. Allen et Amir Husain, auteurs d'un article marquant publié en 2017 dans les *U.S. Naval Institute Proceedings*[2], les progrès exponentiels des capacités de traitement algorithmique ne se contenteront pas de transformer les modalités de la guerre — qu'elle soit conventionnelle ou nucléaire — mais modifieront en profondeur sa nature même. À travers leur concept d'hyper-guerre (*hyperwar*), les auteurs esquissent une rupture fondamentale avec l'héritage clausewitzien qui définissait la guerre comme une dialectique de volontés humaines.

[2] John R. Allen et Amir Husain, « On Hyperwar, » U.S. Naval Institute Proceedings, juillet 2017, Vol. 143, n° 7, https://www.usni.org/magazines/proceedings/2017/july/hyperwar. Voir aussi Amir Husain et al., Hyper-war: Conflict and Competition in the AI Century (Austin: SparkCognition Press, 2018).

Dans cette perspective, la guerre du futur ne serait plus centrée sur la confrontation d'intentions politiques opposées, médiatisée par la force armée, mais dominée par la vélocité autonome de systèmes algorithmiques. La dynamique décisionnelle y serait radicalement accélérée, au point de produire un véritable effondrement des fenêtres temporelles (*collapsing of decision windows*) dans lesquelles une action peut être évaluée, décidée, puis exécutée. Autrement dit, les délais traditionnels de réflexion stratégique et de coordination opérationnelle seraient rendus obsolètes par la vitesse d'action imposée par des agents artificiels autonomes.

Dans un tel environnement, l'avantage reviendrait inévitablement à l'acteur capable de mobiliser le plus grand nombre de systèmes de décision autonomes, tout en déclenchant, sans délai, des réponses ciblées et adaptées contre l'adversaire. La supériorité ne dépendrait plus

Airmen refuel a B-2 Spirit at Whiteman Air Force Base, Mo., May 28, 2025. The 509th Bomb Wing and its fleet of B-2s serve as part of the Air Force's combat force to project U.S. airpower anywhere in the world. @ Air Force Staff Sgt. Joshua Hastings

uniquement de la masse, du feu ou de la manœuvre, mais de la précocité algorithmique — c'est-à-dire de la capacité à percevoir, comprendre et agir avant même que l'ennemi n'ait formulé son propre calcul stratégique.

À l'échelle stratégique, cela suppose une architecture de commandement articulée autour de systèmes intelligents capables de restituer une image synthétique, exhaustive et actualisée en temps réel du théâtre d'opérations. L'IA ne serait plus simplement un outil de soutien, mais un acteur décisif du processus de commandement, en mesure d'orienter ou d'optimiser la prise de décision humaine. C'est cette combinaison d'omniscience opérationnelle et d'ultrarapidité décisionnelle qui, selon Allen et Husain, caractérisera l'hyper-guerre, redéfinissant les équilibres de puissance et les fondements de la stratégie contemporaine.

Qu'il s'agisse des concepts d'hyper-guerre, de guerre algorithmique ou encore de *flash war* — version technologique et anglicisée de la guerre éclair —, ces notions interrogent d'abord la transformation des modalités de la guerre : sa vitesse, sa nature opérationnelle, son degré d'automatisation. Mais une question plus fondamentale se pose désormais : l'IA est-elle susceptible de déstabiliser les fondements mêmes de la dissuasion nucléaire ? Pourrait-elle, en bouleversant les équilibres militaires, précipiter le déclenchement de crises — voire de conflits — qui, sans elle, n'auraient peut-être jamais éclaté ?

C'est à cette interrogation que s'est attachée une étude de la RAND Corporation, fruit de consultations d'experts en IA, de stratèges militaires et de spécialistes des relations internationales. Le cœur de l'hypothèse examinée est le suivant : l'intégration de l'IA dans les architectures de commandement stratégique pourrait-elle, dans certaines configurations, déclencher involontairement un conflit — en particulier nucléaire[3] ?

[3] Edward Geist et Andrew J. Lohn (dir.), How Might AI Affect the Risk of Nuclear War? (Santa Monica: RAND Corporation 2018).

Pour mesurer la gravité d'un tel risque, il faut rappeler que, dès les derniers développements de la guerre froide, tant les États-Unis que l'Union soviétique avaient exploré — et, dans certains cas, déployé — des systèmes automatisés de contrôle des forces nucléaires, censés garantir une capacité de riposte même en cas d'annihilation du commandement humain.

C'est dans ce contexte qu'est conçu, en 1985, le système soviétique Perimetr (connu sous le nom de *Mertvaya Ruka*, ou *Dead Hand*)[4]. Véritable dispositif de dissuasion automatisée, Perimetr visait à assurer une riposte nucléaire massive en toutes circonstances, y compris dans le cas extrême où l'état major aurait été détruit par une première frappe ennemie. Sa logique : pré-enregistrer l'ordre de lancement d'une frappe de représailles, activable automatiquement par un ensemble de senseurs mesurant les effets d'une attaque nucléaire (surpression, rayonnement, perturbations sismiques, silences radio militaires, etc.).

Le système reposait sur deux composants essentiels connus :

- Un missile de commandement (fusée 15P011 équipée d'une ogive radio 15B99), chargé de transmettre l'ordre de riposte aux silos et postes de commandement encore fonctionnels.

- Un système autonome de commandement et de contrôle, équipé de dispositifs de corrélation et d'analyse multisensorielle, capable de déclencher de manière entièrement automatisée le lancement de missiles balistiques intercontinentaux (ICBM) si aucun signal d'interruption humaine n'était détecté.

[4] Petr Topychkanov, "Autonomy in Russian Nuclear Forces," dans The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk: Volume I Euro-Atlantic Perspectives, dir. Vincent Boulanin (Stockholm: SIPRI, 2019), http://www.jstor.org/stable/resrep24525.14; Voir aussi Nicholas Thompson, "Inside the Apocalyptic Soviet Doomsday Machine," Wired, 21 septembre 2009, https://www.wired.com/2009/09/mf-deadhand/.

La logique stratégique derrière Perimetr était claire : opposer à la dissuasion défensive américaine (notamment à travers l'Initiative de défense stratégique – IDS) une posture de dissuasion offensive, inéluctable et déterministe. L'objectif était de démontrer aux États-Unis que toute frappe initiale entraînerait nécessairement une réponse nucléaire soviétique, même en cas de décapitation complète du commandement.

Les États-Unis disposaient d'un système équivalent : l'*AN/DRC-8 Emergency Rocket Communications System* (ERCS). Celui-ci prévoyait le lancement d'un missile *Minuteman II* équipé non pas d'une ogive nucléaire, mais d'un émetteur radio UHF destiné à transmettre les ordres d'alerte du *National Command Authority au Strategic Air Command*, garantissant ainsi la continuité des communications stratégiques en situation extrême.

Ces deux dispositifs montrent que l'idée d'un automatisme dans la décision nucléaire n'est pas nouvelle. Ce qui change aujourd'hui, avec l'IA, c'est la nature des mécanismes décisionnels eux mêmes. Là où Perimetr ou l'ERCS suivaient des logiques prédéfinies, figées, presque mécaniques, les systèmes fondés sur l'IA opèrent dans un environnement dynamique, adaptatif, et parfois opaque, susceptible de produire des raisonnements, des prédictions, voire des recommandations dont l'intelligibilité échappe à l'opérateur humain.

L'introduction de l'IA dans la chaîne décisionnelle stratégique soulève donc un nouveau dilemme de confiance : jusqu'où peut-on déléguer, ou même simplement automatiser, la gestion d'un arsenal nucléaire dans un monde où les cycles décisionnels sont accélérés, les menaces plus ambiguës, et les signaux manipulables ? Ce déplacement de la dissuasion humaine vers une dissuasion potentiellement algorithmique — ou hybride — exige un réexamen en profondeur des doctrines de commandement et des garde-fous éthiques et techniques censés empêcher l'irréparable.

IA ET FEU NUCLEAIRE : LE JUGEMENT HUMAIN EN DERNIER RECOURS

L'idée de déléguer totalement la décision de frappe nucléaire à une machine a toujours suscité de profondes réticences. Ces réserves tiennent au fait que les autorités politiques et militaires ont, jusqu'ici, toujours voulu conserver la prérogative ultime en matière de déclenchement du feu nucléaire. Le système Perimetr illustre parfaitement ces hésitations : bien qu'il fût conçu pour garantir une capacité de riposte automatique en cas d'anéantissement du commandement, il n'en demeurait pas moins subordonné à une décision humaine initiale.

Son surnom de « *Dead Hand* » ne doit donc pas être interprété comme la preuve d'une automatisation intégrale du processus de lancement, mais plutôt comme un mécanisme de continuité. En effet, la plupart des sources s'accordent à reconnaître que l'activation du système – y compris de l'interrupteur dit homme mort – restait à la discrétion d'un opérateur humain. La logique sousjacente n'était pas celle d'un remplacement du jugement humain, mais celle d'une garantie : celle que la dissuasion soviétique pourrait survivre à une décapitation totale, et que la riposte serait malgré tout assurée.

Par ailleurs, les exemples documentés d'erreurs d'évaluation par des systèmes automatisés en matière d'alerte nucléaire abondent dans l'historiographie de la guerre froide. En novembre 1979, un exercice de simulation fut accidentellement chargé dans un ordinateur du NORAD, induisant une fausse alerte de lancement massif de missiles soviétiques. Seule la vérification croisée avec les radars permit de confirmer l'absence d'attaque réelle. Cette procédure, appelée « *dual phenomenology* », consistait à valider toute

alerte nucléaire par au moins deux systèmes indépendants, afin de minimiser le risque de fausse interprétation. Toutefois, une telle redondance n'était pas toujours systématiquement appliquée dans les dispositifs soviétiques.

© *BE Defence*

En juin 1980, un incident encore plus grave manqua de précipiter une réaction catastrophique : Zbigniew Brzeziński, alors conseiller à la sécurité nationale, fut réveillé au milieu de la nuit par un appel l'informant du lancement supposé de 220 missiles soviétiques, chiffre qui grimpa ensuite à 2 200. Sur le point d'alerter le président Jimmy Carter pour ordonner une riposte immédiate, Brzeziński reçut in extremis un troisième appel annulant l'alerte : une puce défectueuse d'une valeur d'un dollar dans l'un des ordinateurs du NORAD était à l'origine de cette défaillance.

En 1983, en Union soviétique, un système d'alerte précoce détecta à tort le lancement de cinq ICBM américains. Le lieutenant-colonel Stanislav Petrov, de service ce jour-là, choisit — contre les procédures — de ne pas valider l'alerte, estimant qu'il s'agissait probablement d'une erreur du système. Il

avait raison : l'algorithme avait pris un reflet solaire sur des nuages pour une signature de lancement de missiles [5]. Sa décision permit d'éviter une possible escalade nucléaire.

Ces épisodes démontrent l'importance irréductible du jugement humain dans la gestion des systèmes de dissuasion, y compris lorsque ceux-ci s'appuient sur des dispositifs automatisés. Ils rappellent aussi que les systèmes technologiques — aussi sophistiqués soient-ils — ne sont jamais exempts d'erreurs, et que la tentation d'une délégation totale à l'IA dans le domaine nucléaire représenterait un saut stratégique aux conséquences potentiellement irréversibles.

Les États-Unis semblent être allés plus loin que d'autres puissances dans l'exploration de l'intégration de l'IA aux procédures liées à la dissuasion nucléaire, en particulier dans des contextes de crise engageant les intérêts vitaux de la nation. Dès la fin des années 1980, le programme SAPE permettait aux forces armées américaines de tester l'utilisation de technologies d'IA pour améliorer les capacités de ciblage des forces stratégiques, notamment à l'encontre des lanceurs mobiles soviétiques d'ICBM. Le système développé dans ce cadre n'avait toutefois aucun contrôle direct sur les armes nucléaires : il s'agissait d'un système expert, destiné à traiter des données de reconnaissance pour élaborer des plans de ciblage dynamiques, transmis ensuite aux bombardiers B-2 chargés de les exécuter.

Ce précédent historique illustre une constante observée au fil des 70 dernières années d'histoire militaire : la question de l'automatisation des fonctions critiques, notamment celles liées à la frappe nucléaire, demeure un enjeu stratégique récurrent au sein des états-majors des puissances technologiquement avancées. L'intégration de systèmes automatisés

[5] John Borrie, « Cold War Lessons for Automation in Nuclear Weapon Systems », dans The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk, Volume I: Euro Atlantic Perspectives, dir. Vincent Boulanin (Stockholm: SIPRI, 2019), 41-52.

ou semi-automatisés dans les chaînes décisionnelles n'a jamais eu pour vocation — du moins selon les sources ouvertes disponibles — de permettre à une machine de déclencher une frappe en premier. L'enjeu était plutôt de préserver la capacité de riposte, notamment dans l'éventualité d'une décapitation du commandement.

Comme l'a souligné Benjamin Hautecouverture, quatre éléments majeurs émergent dans le cadre de cette problématique[6] :

1. Les avancées dans l'automatisation et l'apprentissage machine n'ont pas transformé le cadre stratégique existant. Elles n'ont pas, à ce jour, modifié en profondeur les doctrines ni ouvert de perspectives opérationnelles véritablement inédites. La technologie a évolué, mais les fondements politico-stratégiques de la dissuasion restent inchangés.

2. La méfiance stratégique reste un moteur puissant. Les réflexions sur l'intégration de l'automatisation dans les chaînes de décision s'inscrivent moins dans une logique d'innovation propre que dans une volonté de compréhension des dispositifs adverses. L'exemple du système soviétique Perimetr est révélateur : conçu côté soviétique comme un facteur de stabilité — en évitant d'assigner la décision ultime à un seul dirigeant dans un contexte de chaos — il fut perçu côté occidental comme un élément déstabilisant, opaque et potentiellement incontrôlable.

3. Les limites technologiques ont été récurrentes. Dans chaque tentative d'intégration de l'automatisation à des systèmes de force, les

[6] Benjamin Hautecouverture, « Applications nucléaires de l'automatisation : rappels historiques », in Bulletin mensuel de l'Observatoire de la dissuasion, dir. Bruno Tertrais et Emmanuelle Maître, (Paris : FRS & DGRIS, 2019), 13.

insuffisances techniques sont apparues comme des freins majeurs. Ces limites ont régulièrement donné lieu à un retour au principe de contrôle humain et à la réaffirmation de la nécessité de redondances procédurales — c'est-à-dire la multiplication de points de validation humains et techniques dans les chaînes de décision.

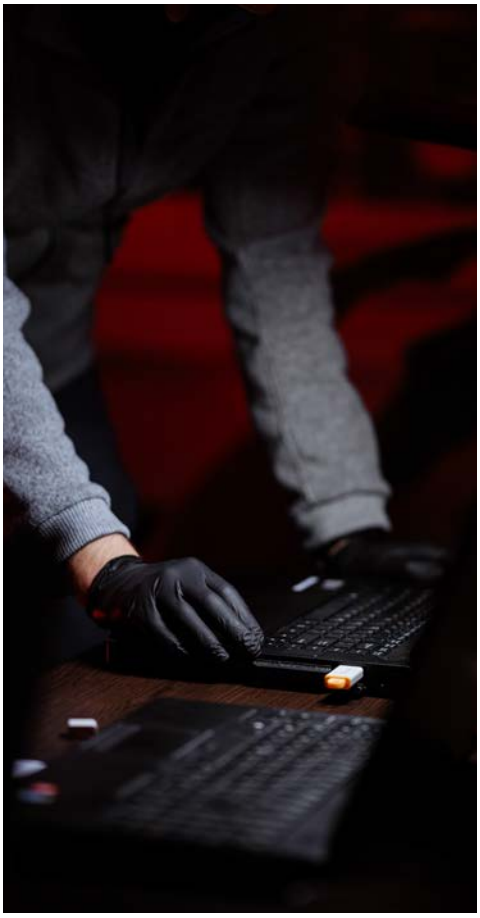
4. La prudence doctrinale a généralement prévalu. De manière générale, les forces armées ayant exploré l'intégration de l'IA dans leurs architectures ont fait preuve d'hésitations structurelles, motivées par des exigences de sûreté et de responsabilité. Il s'agissait d'éviter d'introduire des vulnérabilités systémiques — qu'il s'agisse de failles informatiques, de corruptions de données ou de logiques décisionnelles ininterprétables — dans des systèmes aussi sensibles que ceux impliqués dans la gestion de l'armement nucléaire.

Ces éléments confirment à qui en douterait que, malgré les progrès technologiques spectaculaires, la maîtrise humaine reste perçue comme un principe intangible dans les doctrines nucléaires contemporaines. L'enjeu n'est donc pas tant de savoir si l'IA peut se substituer à l'humain que de déterminer jusqu'où elle peut l'assister sans compromettre la stabilité stratégique globale.

Aujourd'hui, pourtant, de nouvelles incertitudes relatives aux limites des systèmes d'IA apparaissent. Et avec elles leur lot d'interrogations sur notre capacité à préserver les équilibres stratégiques.

BIAIS ET AUTRES DERIVES

Si l'IA est censée offrir aux prévisionnistes ainsi qu'aux organes chargés de la réponse militaire l'opportunité de concevoir des modèles d'anticipation et de



© Quentin Moonen, BE Defence

réaction aux crises plus performants et plus précis, son utilisation n'est pas exempte de risques. Parmi ceux-ci, plusieurs écueils techniques et méthodologiques méritent une attention particulière.

Le premier est bien connu des spécialistes de l'apprentissage automatique : il s'agit du phénomène de surapprentissage (*overfitting*). Ce biais survient lorsque le modèle d'IA, au cours de sa phase d'entraînement, accorde une importance excessive aux moindres détails et aux bruits présents dans les données historiques. En conséquence, le modèle devient trop rigide, incapable de généraliser à partir de données nouvelles ou légèrement différentes. Concrètement, une IA conçue pour analyser en temps réel une multitude de paramètres politico-militaires, et entraînée sur un corpus dense mais figé, risque d'éprouver des difficultés à intégrer efficacement des signaux émergents ou des transformations contextuelles inédites. Cela peut entraîner un retard dans la détection de ruptures ou une mauvaise appréciation de l'évolution d'une situation donnée.

À l'opposé, un second biais mérite d'être évoqué : celui de la surobjectivation des variables issues de sources de renseignement diverses (géospatiales,

humaines, cybernétiques, etc.). Dans ce cas, l'IA pourrait interpréter des fluctuations normales ou des signaux ambigus comme les prémices d'un déséquilibre systémique ou d'une crise imminente. Le risque est alors celui d'une hypersensibilité algorithmique à des perturbations qui, dans un cadre d'analyse humain ou diplomatique, auraient été jugées transitoires ou absorbables par les mécanismes institutionnels en place. En d'autres termes, le modèle pourrait « voir » une crise là où il n'y en a pas, en amplifiant des signaux faibles jusqu'à les interpréter comme des tendances structurantes.

Ces deux effets — surapprentissage et sur-objectivation — illustrent les limites d'un recours non critique à l'IA dans des contextes marqués par une forte complexité stratégique et une dynamique non linéaire. Ils rappellent que, loin de se substituer au jugement humain, l'IA doit être intégrée dans un processus décisionnel hybride, combinant puissance de calcul et discernement analytique.

On peut d'ailleurs se demander ce qu'il adviendrait si ce phénomène de sur-objectivation était sciemment exploité par un adversaire. Voire si cet adversaire — quelle que soit sa nature — parvenait à manipuler intentionnellement les données sources traitées par une IA chargée de la veille stratégique. Cette hypothèse, loin d'être purement spéculative, a fait l'objet de multiples exercices de simulation (*wargames*) visant à évaluer les conséquences d'une manipulation algorithmique sur l'escalade des tensions.[7]

Au cours de ces exercices, impliquant des hauts responsables militaires et civils de pays membres de l'OTAN, plusieurs scénarios ont mis en évidence la vulnérabilité des systèmes d'IA face à des opérations d'influence ciblées. L'usage de *deepfakes*, de contenus audiovisuels falsifiés ou de déclarations fabriquées a ainsi induit en erreur un système d'IA de commandement et de contrôle, le poussant à privilégier des données délibérément inexactes.

[7] Matthew Fitzpatrick, « Artificial intelligence and nuclear command and control, » Survival 61, n° 3 (mai 2019) : 81-92, https://doi.org/10.1080/00396338.2019.1614782.

Plus préoccupant encore, ces expérimentations ont révélé que la fragilisation de l'IA ne se limitait pas à sa phase opérationnelle. Dans certains cas, les données d'entraînement elles-mêmes avaient été altérées en amont — un procédé connu sous le nom de *data poisoning*. Ce sabotage discret avait modifié les pondérations internes du modèle, déformant sa structure analytique de manière durable. L'empoisonnement des données initiales ne conduit donc pas seulement à produire des résultats erronés dans l'analyse du réel : il affecte en profondeur la robustesse de l'algorithme, compromettant la fiabilité stratégique de l'ensemble du dispositif.

Mais encore faut-il être en mesure de détecter de telles altérations avant que l'IA ne produise une analyse situationnelle susceptible d'influencer des décisions critiques en contexte de crise. C'est précisément là que réside l'un des défis majeurs associés à l'utilisation de l'IA dans les sphères politico-militaires : l'opacité intrinsèque des processus qu'elle mobilise — ce que l'on désigne communément par la métaphore de la « boîte noire ».

Le risque fondamental pour les organisations militaires réside dans la perte de capacité à évaluer la validité et la robustesse des raisonnements internes de l'IA. D'où l'importance cruciale des exigences d'explicabilité et de compréhensibilité des systèmes automatisés. Depuis deux décennies, l'essor de l'apprentissage automatique fondé sur des réseaux de neurones profonds (*Deep Neural Networks*, DNN) a bouleversé les rapports entre opérateurs humains et systèmes intelligents. Les corrélations établies par ces algorithmes, aussi précises soient-elles, restent le plus souvent asémantiques — c'est-à-dire dépourvues de signification intelligible pour un humain. Il devient alors extrêmement complexe, voire impossible, de reconstruire une logique causale permettant de justifier les choix, recommandations ou actions proposées par le système.

Cette absence d'interprétabilité soulève une question stratégique majeure : comment accorder une pleine confiance à un système dont les recommandations peuvent avoir un impact direct sur l'identification, la désignation ou même la neutralisation d'une menace, s'il demeure incapable d'expliquer, en des termes compréhensibles, les raisons de ses choix ? Ce paradoxe est d'autant plus préoccupant que les IA les plus performantes en matière de prédiction et d'optimisation stratégique sont aussi, statistiquement, les moins interprétables — échappant ainsi à toute forme de contrôle humain véritablement éclairé.

Ces enjeux prennent une acuité particulière lorsque l'on aborde la question de la dissuasion nucléaire. De nombreux experts en stratégie militaire s'interrogent aujourd'hui sur l'impact de l'IA sur les fondements mêmes de

© *Generated with AI*

la guerre et sur les dynamiques de la dissuasion. L'interrogation centrale pourrait se formuler ainsi : l'intégration croissante de systèmes d'IA au sein des chaînes de commandement et de contrôle (C2) marque-t-elle l'émergence d'un nouveau régime de conflictualité ? Et, si tel est le cas, cette mutation concerne-t-elle uniquement les modalités de la guerre — sa temporalité, sa distribution, ses vecteurs — ou bien sa nature même ?

Ce débat met en lumière un déplacement du centre de gravité des préoccupations stratégiques vers la gouvernance de la guerre dans ses formes contemporaines, qu'elles soient conventionnelles, cybernétiques ou nucléaires. À travers l'IA, c'est bien la manière dont la guerre est pensée, encadrée et conduite qui se transforme — et avec elle, les fondations éthiques, politiques et doctrinales sur lesquelles repose l'usage légitime de la force.

GUERRE, IA ET BOITE NOIRE : VERS UNE PERTE DE CONTROLE STRATEGIQUE ?

Les avancées récentes en IA sont-elles susceptibles de redéfinir en profondeur les cadres stratégiques contemporains — et, plus particulièrement, ceux qui régissent la dissuasion nucléaire ?

Il convient de rappeler que l'automatisation des capacités de frappe nucléaire a constitué, depuis la fin de la Seconde Guerre mondiale, un thème récurrent des débats stratégiques entre puissances nucléaires. Les exemples des systèmes Perimetr et ERCS témoignent d'une volonté historique d'assurer la continuité de la dissuasion, fût-ce par des moyens techniques semi-autonomes. Toutefois, ces dispositifs obéissaient à une logique rigide, fondée sur des conditions bien définies et des seuils mécaniques.

Aujourd'hui, les performances spectaculaires de l'IA, notamment dans le domaine de l'analyse prédictive, du traitement du langage naturel, de la fusion de données multisources ou encore de la reconnaissance automatique de cibles, viennent réactiver ces questionnements dans un cadre technologique profondément renouvelé. La particularité de l'IA contemporaine, en particulier des systèmes d'apprentissage profond, est qu'elle produit des résultats parfois opaques — échappant partiellement à l'intelligibilité humaine — tout en opérant dans des temporalités radicalement raccourcies. Dans un contexte de crise nucléaire, une telle accélération des cycles de décision pourrait remettre en cause les principes de maîtrise, de proportionnalité et de discernement qui sont au fondement même de la dissuasion.

Dès lors, des questions fondamentales restent ouvertes : les progrès actuels de l'IA relèvent-ils d'un développement linéaire et maîtrisable ou s'inscrivent-ils dans une dynamique faite de ruptures technologiques difficiles à anticiper ? Devons-nous redouter l'émergence d'une IA stratégique dotée d'un pouvoir de recommandation si rapide et si opaque qu'il dépasserait les délais de validation humaine ? Ou bien sommes-nous encore, pour l'heure, dans une phase transitoire d'expérimentation, où l'IA ne joue qu'un rôle d'assistance ?

Quelles que soient les réponses à ces questions — qui traduisent autant de postures doctrinales que de visions du monde —, il apparaît que l'impact réel de l'IA sur les systèmes de dissuasion dépendra avant tout du cadre dans lequel elle sera intégrée : cadre juridique, cadre doctrinal, cadre culturel. Car l'IA n'est pas en elle-même une rupture stratégique. C'est l'usage que l'on en fait, et les choix d'architecture politique et militaire qu'elle induit, qui détermineront sa portée transformatrice. L'IA ne redessinera pas les équilibres globaux de manière uniforme ; elle le fera selon les représentations, les intentions et les doctrines de chaque puissance.

CONCLUSION

L'intégration progressive de l'IA dans les dispositifs militaires s'inscrit dans un moment charnière de redéfinition des normes de sécurité internationales. Tandis que s'accélère la transformation des outils de guerre, nous assistons en parallèle à l'érosion des régimes traditionnels de gouvernance stratégique, en particulier dans les domaines nucléaire et cyber. Cette dégradation soulève des interrogations majeures sur la capacité des structures existantes à encadrer l'innovation technologique sans en subir les dérives.

Les défis posés par l'IA ne se limitent pas à l'amélioration des capacités opérationnelles : ils engagent des dimensions juridiques, éthiques et politiques fondamentales, notamment en ce qui concerne la délégation de la décision létale, la responsabilité en cas d'erreur algorithmique ou encore la stabilité de la dissuasion en contexte de confrontation.

Dans ce cadre, il devient impératif de repenser nos mécanismes de contrôle, nos doctrines d'emploi et nos cadres juridiques internationaux. La gouvernance de l'IA dans les affaires militaires ne saurait reposer uniquement sur des approches nationales ou technocentrées. Elle requiert une coordination transnationale, impliquant non seulement les États, mais également les acteurs académiques, les industries de défense et la société civile.

Seule une telle coopération permettra d'anticiper les conséquences systémiques de l'IA sur la paix et la sécurité internationales. Il ne s'agit plus simplement de réguler a posteriori : l'urgence actuelle impose de bâtir des régimes de gouvernance proactifs, capables de suivre le rythme des transformations technologiques et d'en prévenir les effets potentiellement déstabilisateurs.

RECOMMANDATIONS POUR LA BELGIQUE ET L'UNION EUROPÉENNE

Face aux mutations profondes induites par l'intégration de l'IA dans les architectures de commandement stratégique, et plus particulièrement dans les dispositifs liés à la dissuasion nucléaire, la Belgique et l'UE doivent adopter une posture résolument proactive. En tant que membre fondateur de l'UE et de l'OTAN, la Belgique occupe une position charnière au sein des dispositifs de sécurité collective. Sa participation active à la posture de dissuasion élargie de l'Alliance, sa contribution aux programmes du Fonds européen de la défense (FED), ainsi que son implication dans les travaux de gouvernance éthique de l'IA à l'échelle européenne et onusienne lui confèrent une légitimité certaine pour peser dans les débats à venir.

Dans ce contexte, la Belgique devrait avant tout réaffirmer, aux côtés de ses partenaires européens, le principe intangible du contrôle humain dans la chaîne de décision nucléaire. Si les systèmes d'IA peuvent assister les processus d'analyse et de modélisation, ils ne doivent en aucun cas être autorisés à agir de manière autonome lorsqu'il s'agit de décisions engageant l'emploi de l'arme nucléaire. Ce principe du *human-in-the-loop* ou *human-on-the-loop* doit être clairement réaffirmé dans les doctrines de l'UE en matière de technologies émergentes à double usage, mais aussi dans les cadres opérationnels de l'OTAN auxquels la Belgique contribue activement.

Deuxièmement, la Belgique pourrait s'engager, en concertation avec ses partenaires, dans le développement d'un cadre normatif européen sur l'explicabilité et la robustesse des systèmes d'IA utilisés à des fins militaires, notamment dans les fonctions critiques de commandement, de ciblage et d'alerte stratégique. Elle pourrait mobiliser à cette fin les expertises de ses

centres de recherche spécialisés (IRSD, Cyber Command, universités), déjà impliqués dans l'évaluation de projets IA soutenus par le FED, en particulier ceux portant sur la surveillance, la détection de menaces ou la gestion de crise en environnement contesté.

En troisième lieu, la Belgique gagnerait à renforcer ses capacités de résilience algorithmique, notamment contre les risques de *data poisoning*, de manipulation informationnelle et de biais systémiques dans les IA d'aide à la décision. La participation belge à des projets structurants comme SPARTA ou aux initiatives du Centre d'excellence pour la sécurité hybride d'Helsinki donne un levier précieux pour intégrer cette dimension dans les feuilles de route européennes. Elle pourrait également proposer, dans le cadre du prochain programme de travail du FED, un axe de recherche spécifique consacré à la cybersécurité des systèmes algorithmiques en contexte stratégique.

En quatrième lieu, la Belgique devrait soutenir la création d'une instance européenne permanente de réflexion sur l'IA et la dissuasion, réunissant experts civils et militaires, éthiciens et technologues. Cette instance pourrait fonctionner comme un observatoire stratégique et éthique, chargé d'anticiper les effets à moyen et long termes de l'automatisation militaire, tout en formulant des recommandations interinstitutionnelles. Une telle structure pourrait utilement s'inspirer des travaux déjà menés par le SIPRI, l'UNIDIR, la RAND Corporation, ou encore le Partenariat mondial sur l'intelligence artificielle (GPAI), auquel l'UE contribue par l'intermédiaire de ses États membres.

Enfin, sur le plan diplomatique, la Belgique est bien placée pour plaider, au sein des enceintes multilatérales (ONU, CCAC, CD de Genève), pour une initiative européenne sur la régulation des usages de l'IA dans le domaine

de la dissuasion nucléaire. En cohérence avec sa tradition de diplomatie du désarmement et de soutien au multilatéralisme, la Belgique pourrait appeler à l'élaboration d'un code de conduite international sur les systèmes d'IA déployés dans les architectures stratégiques sensibles, à l'image de ce qui a été amorcé pour les systèmes d'armes létaux autonomes (LAWS). Cette démarche contribuerait à éviter l'émergence de zones grises juridiques autour de technologies critiques, tout en consolidant le rôle de l'Europe comme puissance normative dans le domaine des technologies à double usage.

Parce qu'elle n'a nulle raison d'être spectatrice des transformations technologiques à l'œuvre, la Belgique doit se donner les moyens — techniques, diplomatiques, institutionnels — de jouer un rôle moteur dans la définition d'une IA stratégique maîtrisée, éthique et stabilisatrice. L'UE, quant à elle, à l'heure d'un réarmement qui pourrait conduire à une oblitération des principes éthiques, est invitée à se positionner non seulement comme régulatrice, mais également comme actrice stratégique à part entière, capable de structurer un agenda autonome sur les rapports entre IA, sécurité internationale et stabilité nucléaire.

Mots-clés : intelligence artificielle (IA), chaîne décisionnelle nucléaire, dissuasion algorithmisée

An unarmed Minuteman III intercontinental ballistic missile launches during an operational test at Vandenberg Space Force Base, Calif., Nov. 5, 2024. The test launches demonstrate that the U.S. intercontinental ballistic missile fleet is ready, reliable and effective in leveraging dominance in an era of strategic competition.
© Air Force Staff Airman 1st Class Olga Houtsma