

Armes autonomes et cognition étendue

MAJOR ANDREW G. HOLT

Le major Andrew G. Holt est actuellement commandant du 5^e Régiment d'artillerie légère du Canada. Il est titulaire d'une maîtrise en histoire et philosophie des sciences et de la technologie de l'Université de Toronto et d'une maîtrise en études militaires de l'Université Deakin.

Les pays se développent rapidement et mettent en œuvre des SALA. Ces armes font l'objet d'un débat dans les milieux universitaire, juridique et politique, où beaucoup soulèvent des objections éthiques en raison de lacunes perçues sur le plan de la responsabilisation. De telles lacunes pourraient entraver les poursuites contre les criminels de guerre, menacer la conduite morale de la guerre et réduire le respect du droit des conflits armés.¹ Beaucoup se concentrent sur le maintien des SALA sous contrôle humain, mais sans consensus sur la façon dont ce contrôle est défini ainsi que sur la façon dont il peut être exercé. La HCE, une vision de plus en plus populaire dans la science cognitive contemporaine, offre un moyen utile de comprendre le contrôle humain des SALA. J'aborde en premier lieu le débat en cours sur les SALA et la difficulté à définir le contrôle humain. Je résume ensuite la HCE et son application au contrôle humain des SALA, en examinant la façon dont les humains interagissent avec les technologies cognitives en général, en considérant les SALA comme des artéfacts cognitifs et en comparant le contrôle des SALA et des subordonnés humains. Je conclus que les SALA peuvent étendre la cognition et être sous le contrôle humain, même sans intervention humaine immédiate et consciente. Par conséquent, les processus militaires tels que l'instruction et la diffusion d'ordres pourraient permettre aux commandants d'exercer le contrôle des SALA de manière parallèle au contrôle des subordonnés humains.

Notions vagues de contrôle

Les universitaires débattent des répercussions éthiques des SALA depuis plus de dix ans. À la lumière de ce débat, la plupart d'entre eux adoptent des positions sceptiques à l'égard des SALA, sinon carrément opposées à leur utilisation. Une préoccupation commune est le manque de responsabilité humaine pour les SALA.² Une minorité d'auteurs, comme Simpson et Muller, ont un point de vue différent.³ Ils soutiennent que la responsabilité incombe aux commandants qui ont armé et déployé les SALA, et que le fait que les SALA se comportent de manière imprévisible n'empêche pas leur utilisation. Le débat n'est pas seulement théorique, mais mène à des appels à l'interdiction ou à la limitation des SALA en raison d'un manque de contrôle humain.⁴

Diverses organisations non gouvernementales ont adopté des positions fermes sur les SALA. L'International Committee for Robot Arms Control préconise que nous devons agir contre « les dangers pressants que [les SALA] posent à la paix et à la sécurité internationale ainsi qu'aux civils en temps de guerre⁵ » [TCO]. Le Comité international de la Croix-Rouge (CICR) a fait une présentation à une réunion des Hautes Parties contractantes et signataires de la Convention sur certaines armes classiques (CCAC) axée sur le contrôle humain des SALA, déclarant que : « il est essentiel que les États développent une compréhension commune des éléments

et critères du contrôle humain⁶ » [TCO]. Même les organisations qui s'intéressent de loin aux SALA, comme l'Institute of Electrical and Electronics Engineers (IEEE), recommandent que des mesures soient prises pour « aider à assurer un contrôle humain significatif des systèmes d'armes⁷ » [TCO].

Des États ont également commenté les répercussions juridiques et éthiques des SALA, bien que les positions énoncées puissent ne pas être exhaustives ni harmonisées avec les actions réelles desdits États.⁸ De nombreuses positions articulées se concentrent sur le contrôle humain des SALA, mais il n'y a pas de consensus sur ce que signifie réellement le contrôle humain.⁹ Le président du Groupe d'experts gouvernementaux a proposé une caractérisation de travail des SALA en 2025, mais le débat qui s'ensuit montre que la question du contrôle humain n'est toujours pas résolue. Certaines parties ont quand même demandé la qualification de « contrôle humain » avec des expressions telles que « contexte approprié » ou « significatif », mais il n'y a pas eu d'accord, car « les délégations ont réitéré leurs diverses positions partagées au cours des [précédents] échanges¹⁰ » [TCO].

Les discussions sur le contrôle humain font souvent référence à un cycle décision-action ou encore à l'exigence de la surveillance ou de l'intervention humaine. Le Canada utilise l'exigence de la surveillance ou de l'intervention humaine dans les documents officiels sur les SALA, en disant dans la stratégie

de modernisation de l'Armée canadienne que « l'exigence de la surveillance ou de l'intervention humaine demeurera pour toute application de la force létale¹¹ ». En 2013, Human Rights Watch a formulé les définitions suivantes :

- Armes semi-autonomes : Des robots qui peuvent sélectionner des cibles et fournir une force uniquement avec un commandement humain;
- Armes autonomes supervisées : Des robots qui peuvent sélectionner des cibles et fournir une force sous la supervision d'un humain qui peut passer outre les actions des robots;
- Armes autonomes : Des robots capables de sélectionner des cibles et de fournir une force sans intervention ou interaction humaine.¹²

Qu'est-ce qui constitue un « commandement humain » ou une « surveillance » en termes de temps, d'espace et de spécificité? Qu'en est-il d'une arme autonome « supervisée » pour rechercher et détruire la plus grande cible dans une zone donnée? La surveillance humaine doit-elle être constante et en temps réel? Tenter de définir le contrôle humain en faisant appel à des concepts comme « commandement » et « surveillance » ne fait que pousser le fardeau explicatif vers de nouveaux termes. Certains spécialistes soutiennent que le débat entourant les SALA demeure « truffé d'un manque de connaissances techniques et de rhétorique affective¹³ ». De plus, la surveillance humaine risque de produire un « biais d'automatisation » qui, selon Cummings, peut mener à de nouvelles erreurs si les capacités cognitives humaines ne sont pas prises en compte.¹⁴

La notion mal définie de contrôle humain entrave les progrès dans le débat sur les plans juridique et éthique sur les SALA. Le CICR note qu'« il y a un accord général entre les États sur le fait que le contrôle humain—qu'il soit qualifié de “significatif”, “substantiel”, “approprié”, “suffisant” ou “efficace”, ou “à des niveaux appropriés de jugement humain” ou “d'implication humaine”—doit être conservé par rapport aux systèmes d'armes et au recours à la force¹⁵ » [TCO]. Cette prolifération d'adjectifs révèle un manque de consensus et de clarté. Bode et Heulss font valoir que « [l]a compréhension multiple de ce qu'implique un contrôle humain significatif présente une complication¹⁶ » [TCO]. Les pays favorables à l'élaboration de SALA se demandent donc si et comment le contrôle humain peut être défini de manière utile, soutenant qu'« il est douteux que des critères permettant de déterminer un niveau d'importance attendu du contrôle humain sur la machine puissent être élaborés¹⁷ » [TCO]. Les mesures réglementaires sembleraient impossibles à mettre en œuvre sur la base d'une notion vague de contrôle, sans parler des mesures qui peuvent être appliquées de façon pratique dans le cadre d'opérations militaires.

Pour compliquer davantage le contrôle humain, les SALA peuvent être autonomes de différentes manières et à des degrés différents.¹⁸ Brian Cox, officier à la retraite de la U.S. Army,

propose que nous comprenions les SALA en termes de degré d'autonomie, de ce que cette autonomie apporte à leur fonction générale et de leur méthode de programmation.¹⁹ Même en 2012, avant de nombreuses avancées en matière d'IA, le département de la Défense des États-Unis a fait remarquer que « l'autonomie du système est un continuum allant du contrôle humain complet de toutes les décisions à des situations où de nombreuses fonctions sont déléguées à l'ordinateur²⁰ » [TCO]. Dans un rapport publié en 2017, les auteurs explorent les définitions de l'autonomie et décrivent l'autonomie des armes en termes de capacité accomplie de manière autonome.²¹ Ils constatent que divers systèmes (défense aérienne, protection active, munitions rôdeuses, etc.) peuvent être autonomes dans diverses fonctions (p. ex. mobilité, ciblage et gestion de la santé²²).

Il est utile, compte tenu de la grande variété des SALA, de concentrer mon analyse sur un type particulier. Les systèmes aériens sans pilote (UAS) ou « drones » qui sont autonomes dans la sélection et la poursuite de cibles sont les « robots tueurs » archétypaux au centre du débat des SALA. La discussion de ces systèmes n'est pas seulement hypothétique – un rapport des Nations Unies indique une première utilisation possible de ces SALA en Libye en mars 2020.²³ L'Ukraine a beaucoup utilisé les UAS dans la guerre en cours avec la Russie.²⁴ Certains effectuent des attaques de manière autonome.²⁵ C'est ce type de système, avec ce haut degré d'autonomie dans le ciblage, que j'examinerai en explorant la façon dont le contrôle humain des SALA pourrait être compris grâce à la HCE.

Hypothèse de la cognition étendue (HCE)

Les sciences cognitives et la philosophie de la cognition offrent un moyen utile de comprendre le contrôle humain significatif des SALA. L'examen de la manière dont les SALA pourraient étendre la cognition recoupe un ensemble de réflexions existantes sur l'IA, les artefacts cognitifs et la HCE. Les SALA constituent un exemple particulièrement intéressant et éthique de cognition étendue qui semble encore inexploré. Aucune des publications existantes sur les SALA ne tient compte des SALA par l'entremise de la HCE. Suchman et Weber font une référence intrigante à l'ouvrage fondateur de Knorr-Cetina, *Epistemic Cultures*, et à ses implications pour les relations sujet/objet,²⁶ mais n'explorent pas davantage la façon dont les SALA et les soldats pourraient faire partie d'un système cognitif.²⁷ C'est la lacune que j'essaie de combler. Avant d'examiner les SALA en particulier, il sera utile de donner un aperçu de la HCE.

La HCE possède un héritage intellectuel qui remonte à plusieurs décennies. Clark et Chalmers soutiennent que l'esprit humain peut s'étendre au monde physique et matériel, et qu'il le fait effectivement.²⁸ Leur théorie repose sur la parité fonctionnelle entre les rôles des processus internes (neuraux) et externes dans la cognition. Ils s'opposent aux tentatives visant à limiter l'esprit humain en se basant uniquement sur le lieu où se

déroule un processus cognitif. Un autre argument en faveur de la théorie de l'esprit étendu est que nos outils complètent souvent les processus cognitifs du cerveau, plutôt que de les imiter ou de les reproduire.²⁹ Ces deux arguments arrivent à la même conclusion : l'esprit ne réside pas seulement dans le cerveau. Les mécanismes physiques qui permettent nos processus cognitifs et, en fait, notre *esprit*, ne se limitent pas à nos processus neuronaux. Notre esprit peut parfois s'étendre aux outils que nous utilisons, nos cahiers de notes, télescopes, marteaux, ordinateurs portables et ainsi de suite. Bien que la HCE reste controversée, elle gagne en popularité dans les sciences cognitives. Comme le note Heersmink : « Le changement de paradigme, passant d'une conceptualisation de l'esprit en tant qu'affaire purement interne à la perception qu'il est incarné et façonné par l'environnement et fortement lié à celui-ci, est bien amorcé³⁰. » [TCO]. Je ne tente pas d'entrer pleinement dans le débat sur les HCE; il suffit d'avoir établi que la HCE est une théorie crédible.

Dans le cadre du paradigme de la HCE, les SALA pourraient être considérés comme des artefacts pouvant étendre la cognition humaine ou des artefacts cognitifs. Heersmink appelle les artefacts cognitifs « objets ou structures matériels d'origine humaine qui contribuent de manière fonctionnelle à l'exécution d'une tâche cognitive³¹ » [TCO]. Des auteurs tels que Sienknecht ont envisagé l'intégration d'artefacts cognitifs dans la guerre, où ces artefacts fournissent des intrants aux processus cognitifs biologiques (comme un système d'IA qui analyse les mouvements de l'ennemi et recommande à un commandant humain une frappe préventive³²). Un SALA permettrait plutôt d'étendre la cognition biologique à l'action. Si l'esprit s'étend dans les SALA, comme le permet la HCE, peut-être que cette extension reste sous contrôle humain.

Les SALA pour étendre la cognition

Les SALA, en tant qu'artefacts cognitifs, peuvent assumer certaines des tâches cognitives requises pour réaliser les intentions de leurs concepteurs et opérateurs. Pour comprendre comment, nous devons examiner les intentions plus en détail. Les intentions peuvent être hiérarchiques et exister à différents niveaux d'abstraction. Un commandant humain peut déployer des SALA avec l'intention plus abstraite, ou supérieure, de perturber les communications ennemies. Un SALA pourrait analyser le trafic radio ennemi, construire une représentation du réseau de communication ennemi ainsi que localiser et détruire les nœuds clés qui s'y trouvent. Il pourrait le faire en ciblant les radios elles-mêmes, les antennes de retransmission ou peut-être le carburant qui alimente leurs génératrices. Tout en fonctionnant de manière autonome en termes de sélection et de poursuite de cibles particulières, un tel SALA continue d'exercer une intention humaine. Les SALA permettent simplement aux humains de concentrer leurs intentions et pensées conscientes à un niveau plus élevé d'abstraction. C'est exactement le genre de délégation complémentaire des tâches cognitives qui, selon Heersmink,

caractérise un système cognitif étendu : « Le cerveau humain se développe dans un environnement riche en informations et doit apprendre à compléter ses lacunes en utilisant les fonctionnalités des artefacts pour effectuer des tâches cognitives de manière à ne pas gaspiller des ressources internes coûteuses³³. » [TCO]. Hernandez-Orallo et Vold parlent de technologies qui agissent comme des extensions de l'IA, qui existent « dans un continuum entre des systèmes complètement extériorisés, faiblement couplés à des humains, et des processus entièrement internalisés³⁴ » [TCO]. Les SALA semblent se situer sur ce spectre, même en l'absence d'un humain.

Une propriété intéressante des SALA avancés serait leur transparence³⁵ envers le commandant qui les emploie. Ce manque de conscience peut même justifier la *nécessité* de comprendre les SALA comme des extensions cognitives. Commentaires de Heersmink : « L'utilisation des artefacts peut être conceptualisée selon un spectre d'expertise. La première fois que nous utilisons un artefact, cela nécessite beaucoup d'efforts conscients et cognitifs, mais avec le temps, nous devenons plus aptes à l'utiliser et son utilisation devient plus procédurale³⁶. » [TCO]. Le philosophe Michael Wheeler soutient qu'une telle transparence peut être une indication de l'extension cognitive, mais aussi un motif de préoccupation.³⁷ Si les systèmes qui étendent notre cognition deviennent suffisamment transparents pour nous, nous pourrions ne plus réaliser que leurs catégorisations du monde ne correspondent pas aux nôtres.³⁸ Wheeler écrit :

« Si une application de réseau neuronal profond à laquelle je suis couplé de manière transparente se qualifie comme faisant partie de mon architecture cognitive et donc comme faisant partie de moi, alors les classifications en question—qui guident inconsciemment mon comportement—feront partie de ce que je crois inconsciemment être le cas, et aura donc vraisemblablement le même statut que mes croyances inconscientes plus familières, réalisées en interne³⁹. » [TCO]

Son objection est que ces machines « peuvent classifier régulièrement le monde d'une manière potentiellement dommageable et épistémique par rapport à nos pratiques technologiquement non augmentées⁴⁰ » [TCO]. Cette situation est particulièrement préoccupante pour les SALA qui utilisent l'IA et fonctionnent comme des « boîtes noires » dont le fonctionnement interne est inconnu de l'utilisateur. Il recommande donc que les technologies cognitives soient délibérément non transparentes, comme les appels à garder l'intervention d'un humain pour les SALA.

Les préoccupations au sujet des technologies cognitives ne sont pas nouvelles. Socrate s'inquiète du fait que l'écriture, en tant que technologie cognitive, agit en dehors de tout contrôle humain significatif, affirmant : « Et une fois qu'une chose est mise par écrit, la composition, quelle qu'elle soit, dérive dans tous les sens, tombant entre les mains non seulement de ceux qui la comprennent, mais aussi de ceux qui n'ont rien à voir avec elle; il ne sait pas comment s'adresser aux bonnes personnes, et ne sait pas s'adresser aux mauvaises⁴¹. » [TCO]. Ces mots font

écho aux objections modernes à l'égard des SALA qui interagissent involontairement avec les mauvaises personnes. Si nous acceptons que les intentions d'un auteur sont suffisamment intégrées dans une composition écrite pour qu'il n'ait pas besoin de contrôle humain actif lorsqu'il se propage sans supervision dans le monde, il devrait y avoir une certaine parité dans le degré et le type de contrôle que nous exigeons sur les SALA. Bien que les effets potentiellement mortels des SALA puissent limiter la comparaison directe avec la technologie civile, les SALA ne sont pas la seule technologie dont la défaillance présente un risque de décès (c.-à-d. ponts et automobiles).

La HCE, en permettant que l'esprit puisse s'étendre dans les SALA, remet en question toute insistance sur le fait qu'un cerveau doit superviser le contrôle. De plus en plus, l'apprentissage machine et l'IA permettront aux SALA d'avoir un effet non seulement sur une intention humaine spécifique, en assumant une infime partie de la cognition, mais aussi sur une intention humaine générale ou plus abstraite en assumant un fardeau beaucoup plus cognitif. Penser au contrôle simplement en termes d'implication immédiate et active d'un cerveau humain ne tient pas compte de la façon dont nous étendons systématiquement la cognition aux artefacts physiques sans céder le contrôle. Le CICR soutient qu'« un système d'armes doté d'une autonomie dans ses fonctions critiques *empêche le commandant ou l'opérateur humain de porter les jugements requis par le DIH [droit international humanitaire⁴²] [italiques ajoutés]* » [TCO]. Cependant, la HCE permet que l'esprit qui porte ces jugements *s'étende* aux SALA. Dans le cadre de la HCE, une force militaire pourrait instaurer des processus cognitifs soit dans un cerveau humain ou dans le monde physique, ou un mélange des deux.

La question de savoir si un SALA réalise l'intention humaine peut être plus importante que le fait qu'elle soit sous supervision humaine directe et immédiate. La présentation de 2018 des États-Unis à la CCAC fait avancer cette idée en énonçant ce qui suit : « La question clé de l'interaction homme-machine dans le développement, le déploiement et l'utilisation des technologies émergentes dans le domaine des systèmes d'armes autonomes létales est de s'assurer que, lorsqu'il est nécessaire de recourir à la force, cette force est utilisée pour réaliser les intentions des commandants et des opérateurs⁴³. » [TCO]. L'Australie adopte un point de vue similaire, déclarant que « l'expression "contrôle humain" utilisée par certains lors des discussions [du Groupe d'experts gouvernementaux] ne couvre pas adéquatement la multitude de contrôles pratiques ou de systèmes utilisés par les États et leurs forces armées⁴⁴ » [TCO]. Plutôt que de mandater la supervision par l'être humain, l'Australie articule plutôt son propre système de contrôle en neuf étapes applicable à toutes les armes, commandées par des humains et autonomes.

Conséquences pour les commandants

L'autonomie des subordonnés humains n'implique pas qu'ils soient hors du contrôle de leurs commandants. Les armées n'ont

pas besoin d'un commandant qui surveille un subordonné humain et autorise chaque action spécifique. James Kraska, du U.S. Naval War College, soutient qu'il n'y a pas d'écart de responsabilité avec les SALA, car les commandants demeurent ultimement responsables de tous les moyens de guerre qu'ils utilisent pour atteindre leurs objectifs militaires.⁴⁵ Certes, les SALA et les subordonnés humains ne sont pas identiques, les subordonnés humains ayant une responsabilité et un pouvoir moral que les SALA n'ont pas.⁴⁶ Malgré cette disparité, le contrôle des subordonnés humains peut encore influencer le contrôle des SALA à au moins deux égards : les ordres et l'instruction.

Les commandants donnent des ordres qui dirigent les actions de leurs subordonnés humains. Ces ordres peuvent comprendre à la fois une tâche et un but. Les ordres font généralement référence à l'intention du commandant supérieur, y compris des informations contextuelles qui permettent aux subordonnés humains d'agir conformément à ces intentions, même si la situation change si radicalement que leur tâche assignée n'est plus appropriée. Les intentions d'un commandant peuvent être exécutées de manière opaque pour ce dernier sans perte de contrôle. Même dans les armées modernes et professionnelles avec une communication avancée, il existe des exemples où des subordonnés ont agi de manière inattendue, sans supervision par les supérieurs, pour atteindre une intention supérieure.⁴⁷ Ces unités militaires ne sont pas considérées comme hors de contrôle, alors pourquoi les SALA devraient-ils l'être? Le cas échéant, les ordres peuvent contenir une variété de contraintes officielles (règles d'engagement, limites, échéanciers et exigences en matière de rapports) qui pourraient avoir des analogies directes pour les SALA. Il existe un nombre croissant de documents qui cherchent à « évaluer la fiabilité des humains et des systèmes automatisés pour décider s'il faut se fier à un mandataire donné pour effectuer une certaine tâche⁴⁸ » [TCO]. Un commandant humain peut avoir une confiance instrumentale suffisante dans les SALA pour appliquer une forme de commandement de mission et accepter des actions imprévues qui atteignent toujours l'intention.

L'analogie avec l'entraînement des humains fournit un moyen d'observer l'interaction entre les humains et les SALA et le contrôle des SALA utilisant l'IA ou l'apprentissage automatique. Les SALA qui utilisent des modèles génératifs sont « entraînés » sur un échantillon d'intrants simulant ce qu'ils pourraient rencontrer dans les opérations. Cette formation offrirait des possibilités de confirmation et de validation des SALA très semblables à ce qui est fait pour les unités militaires (humaines) avant le déploiement. Bien que Verbruggen soutienne qu'« il n'existe aucune technique permettant de valider et de vérifier l'IA, pour s'assurer que les systèmes fonctionneront comme prévu et éviteront les comportements imprévisibles⁴⁹ » [TCO]. Cette compréhension de la validation, qui vise une prévisibilité totale, est beaucoup plus stricte que ce qui est appliqué aux organisations humaines. L'Armée canadienne définit la confirmation comme « une évaluation qui montre à quel point une

entité tactique a satisfait à une [norme relative aux tâches de combat⁵⁰]. La validation sert à « vérifier que le programme d'[instruction collective] a bien préparé les unités ou les formations à remplir des tâches opérationnelles permettant d'atteindre des objectifs précis de l'Armée canadienne⁵¹ ». Les deux sont une responsabilité des commandants.

Un SALA bien entraîné peut être sous contrôle même s'il fonctionne en dehors de la conscience humaine et de la prise de décision, comme l'exécution inconsciente d'exercices humains. Dans certaines disciplines militaires, les réactions et les comportements nécessitant une réponse immédiate sont entraînés jusqu'à un point de quasi-inconscience. Avec un entraînement suffisant, cette manipulation d'une arme pour des fins potentiellement létales devient presque transparente. Wheeler décrit cela comme la transparence de l'utilisation des outils, en écrivant :

« Lorsque nous manipulons habilement de l'équipement sans accroc, nous n'avons aucune appréhension consciente des articles d'équipement utilisés en tant qu'objets indépendants, c'est-à-dire quelque chose comme des porteurs identifiables d'états et de propriétés déterminés. Ainsi, pour reprendre l'exemple le plus cité de Heidegger, pendant qu'il se livrait à un martelage sans problème, le menuisier qualifié n'a pas la reconnaissance consciente du marteau, des clous ou du banc de travail, comme on le ferait si on prenait du recul et y réfléchissait⁵². » [TCO]

Une arme peut être semblable à un marteau de menuisier habile. Étant libérée de la nécessité de se concentrer consciemment sur la façon de manipuler l'arme, la conscience humaine peut se concentrer sur des tâches cognitives plus abstraites, comme l'effet que cette manipulation vise à obtenir (c.-à-d. la suppression d'un tireur d'élite ennemi). Si l'entraînement d'un humain à répondre inconsciemment n'entraîne pas une abdication du contrôle significatif sur la force létale, la programmation ou l'entraînement d'un SALA artificiellement intelligent

pourrait-il également fournir le contrôle, même si les SALA opèrent en dehors de la conscience consciente humaine? Écarté des décisions individuelles concernant la sélection et la poursuite des cibles, un commandant militaire n'est pas détaché de l'usage de la force ni de ses responsabilités, mais il est plutôt libre de l'envisager à un niveau supérieur.

Conclusion

La notion de contrôle humain des SALA est complexe, et son examen par la HCE peut mener à des idées et des politiques utiles. Traiter les SALA comme des artéfacts cognitifs ouvre de nouvelles façons de comprendre et d'exercer le contrôle humain des SALA. Les SALA qui étendent la cognition humaine peuvent être sous contrôle même sans intervention immédiate et consciente par un cerveau humain. L'ouverture à l'expérimentation et à la formation avec les SALA pourrait produire des avantages imprévus pour les FAC. Carter et Nielsen soutiennent que les modèles génératifs peuvent permettre une augmentation de l'intelligence artificielle, écrivant ceci :

« [Les modèles génératifs] peuvent être utilisés pour explorer et découvrir, fournir de nouvelles représentations et opérations, qui peuvent être intériorisées dans le cadre de la pensée de l'utilisateur. Et bien que ces exemples en soient à leurs débuts, ils suggèrent que l'IA n'est pas seulement une externalisation cognitive. Une vision différente de l'IA est possible, une vision où elle nous aide à inventer de nouvelles technologies cognitives qui transforment notre façon de penser⁵³. » [TCO]

Si les SALA sont entraînés en collaboration avec leurs homologues humains, les solutions qu'ils emploient peuvent en fait fournir de nouvelles idées et de nouveaux outils cognitifs qui élargissent l'éventail de la pensée humaine. Cette valeur ne devrait pas être modifiée.

Notes

- 1 Mariarosaria Taddeo, *The Ethics of Artificial Intelligence in Defence*, New York, Oxford University Press, 2024, p. 181.
- 2 Robert Sparrow, « Killer Robots », *Journal of Applied Philosophy* 24, n° 1 (2007), p. 62 à 77.
- 3 Thomas Simpson et Vincent Müller, « Just War and Robots' Killings », *Philosophical Quarterly* 66, n° 263 (2016), p. 302 à 322.
- 4 Ingvild Bode et Hendrik Huelss, *Autonomous Weapons Systems and International Norms*, Kingston, McGill-Queen's University Press, 2022, p. 9.
- 5 International Committee for Robot Arms Control, consulté le 3 novembre 2025. Source originale : <https://www.icrac.net/about-icrac/>
- 6 Comité international de la Croix-Rouge, « The Element of Human Control », *Meeting of the High Contracting*

- Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, 2018, p. 1. Source originale : https://reachingcriticalwill.org/images/documents/Disalibr-fora/ccw/2018/hcp-meeting/documents/ICRC_WP3.pdf, consulté le 3 novembre 2025.
- 7 IEEE, « Reframing Autonomous Weapons Systems », *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*, p. 113. Source originale : https://standards.ieee.org/content/dam/ieee-standards/standards/web/other/lead_reframing_autonomous_weapons_v2.pdf, consulté le 3 novembre 2025.
- 8 Michael Schmitt et Sean Watts, « The Decline of International Humanitarian Law Opinio Juris and the Law of Cyber Warfare », *Texas International Law Journal* 50, n° 2-3 (2015), p. 191.

- 9 Australie, « Australia's System of Control and Applications for Autonomous Weapon », *Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*, 2018. Source originale : <https://reachingcriticalwill.org/images/documents/Disalibr-fora/ccw/2019/gge/Documents/GGE.2-WP2.pdf>, consulté le 3 novembre 2025.
- 10 Group of Governmental Experts on Emerging Technologies, « Chair's summary - First 2025 session of the GGE on LAWS », *Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*,

- 2025, p. 7. Source originale : [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2025\)/CCW-GGE.I-2025-WP.I.-Chair's_summary.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2025)/CCW-GGE.I-2025-WP.I.-Chair's_summary.pdf), consulté le 6 octobre 2025.
- 11 Quartier général de l'Armée canadienne, *Engagés vers l'avant : la stratégie de modernisation de l'Armée canadienne*, 2020, p. 54. Source originale : https://publications.gc.ca/collections/collection_2021/mdn-dnd/D2-438-2021-fra.pdf, consulté le 3 novembre 2025.
 - 12 Human Rights Watch, *Losing Humanity, The Case Against Killer Robots*, 2012. Source originale : <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots>, consulté le 3 novembre 2025. Bien qu'elle soit désuète, cette typologie demeure influente et a été utilisée par Vincent Boulanin et Maakie Verbruggen dans *Mapping the Development of Autonomy in Weapon Systems*, Stockholm, Stockholm International Peace Research Institute, 2017, p. 26, pour caractériser les systèmes dans leur ensemble de données.
 - 13 Mary Cummings, « Lethal Autonomous Weapons : Meaningful Human Control or Meaningful Human Certification? [Opinion] », *IEEE Technology and Society Magazine* 38, n° 4 (1^{er} décembre 2019), p. 20-26. Source originale : <https://technologyandsociety.org/lethal-autonomous-weapons-meaningful-human-control-or-meaningful-human-certification/>, consulté le 28 février 2025.
 - 14 Missy Cummings, « Automation Bias in Intelligent Time Critical Decision Support Systems », *American Institute of Aeronautics and Astronautics 1st Intelligent Systems Technical Conference*, du 20 au 22 septembre 2004. Source originale : <https://arc.aiaa.org/doi/10.2514/6.2004-6313>, consulté le 9 juillet 2025.
 - 15 Comité international de la Croix-Rouge, « The Element of Human Control », *Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*, 2018, p. 1. Source originale : https://reachingcriticalwill.org/images/documents/Disalibr-fora/ccw/2018/hcp-meeting/documents/ICRC_WP3.pdf, consulté le 6 octobre 2025.
 - 16 Ingvild Bode et Hendrik Huels, *Autonomous Weapons Systems and International Norms*, Kingston, McGill-Queen's University Press, 2022, p. 159.
 - 17 Fédération de Russie, « Russia's Approaches to the Elaboration of a Working Definition and Basic Functions of Lethal Autonomous Weapons Systems in the Context of the Purposes and Objectives of the Convention », *Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*, 2018, p. 3. Source originale : <https://reachingcriticalwill.org/images/documents/Disalibr-fora/ccw/2018/gge/documents/GGE.I-WP6-English.pdf>, consulté le 6 octobre 2025.
 - 18 Noel Sharkey, « Staying in the Loop : Human Supervisory Control of Weapons », dans *Autonomous Weapons Systems: Law, Ethics Policy*, sous la dir. de N. Bhuta, S. Beck, R. Geiß, H.Y. Liu et C. Kreß, Cambridge, Cambridge University Press, 2016, p. 27.
 - 19 Brian Cox, « Why Binding Limitations on Autonomous Weapons Will Remain Elusive », *Articles of War*, 2022. Source originale : <https://lieber.westpoint.edu/binding-limitations-autonomous-weapons-elusive>, consulté le 6 avril 2022, exposé plus en détail par l'auteur lors d'une présentation à The Canadian National Group of the International Society for Military Law and the Law of War le 24 février 2022, à Ottawa (diffusion virtuelle sur MS Teams).
 - 20 DoD Defense Science Board, *The Role of Autonomy in DoD Systems*, 2012, p. 4. Source originale : <https://apps.dtic.mil/sti/pdfs/ADA566864.pdf>, consulté le 28 février 2025.
 - 21 Vincent Boulanin et Maakie Verbruggen, *Mapping the Development of Autonomy in Weapon Systems*, Stockholm, Stockholm International Peace Research Institute, 2024, p. 20.
 - 22 Vincent Boulanin et Maakie Verbruggen, *Mapping the Development of Autonomy in Weapon Systems*, Stockholm, Stockholm International Peace Research Institute, 2024, p. 36.
 - 23 Nations Unies, *Lettre datée du 8 mars 2021, adressée à la Présidente du Conseil de sécurité par le Groupe d'experts sur la Libye créé par la résolution 1973 (2011) du Conseil de sécurité*, 2021, p. 20. Source originale : <https://docs.un.org/fr/s/2021/229>, consulté le 25 octobre 2025.
 - 24 Jon-Wyatt Matlack, Sebastien Schwartz et Oliver Gill, *Ukraine's Drone Ecosystem and the Defence of Europe: Lessons Lost Can't Be Learned*, Londres, LSE, 2025. Source originale : <https://www.forbes.com/sites/davidhambling/2023/10/17/ukraines-ai-drones-seeking-and-attack-russian-forces-without-human-oversight/>, consulté le 21 septembre 2024.
 - 25 David Hambling, « Ukraine's AI Drones Seek and Attack Russian Forces Without Human Oversight », *Forbes*, 17 octobre 2023. Source originale : <https://www.forbes.com/sites/davidhambling/2023/10/17/ukraines-ai-drones-seeking-and-attack-russian-forces-without-human-oversight/>, consulté le 21 septembre 2024.
 - 26 Lucy Suchman et Jutta Weber, « Human-Machine Autonomies », dans *Autonomous Weapons Systems: Law, Ethics Policy*, sous la dir. de N. Bhuta, S. Beck, R. Geiß, H.Y. Liu et C. Kreß, Cambridge, Cambridge University Press, 2016, p. 98.
 - 27 Karin Knorr-Cetina, *Epistemic Cultures : How the Sciences Make Knowledge*, Cambridge, Harvard University Press, 1999.
 - 28 Andy Clark et David Chalmers, « The Extended Mind », *Analysis* 58, 1998, p. 7-19.
 - 29 Richard Heersmink, « Dimensions of Integration in Embedded and Extended Cognitive Systems », *Phenomenology and the Cognitive Sciences* 14, 2015, p. 577 à 598.
 - 30 Richard Heersmink, « Varieties of Artifacts: Embodied, Perceptual, Cognitive, and Affective », *Topics in Cognitive Science* 13, 2021, p. 574.
 - 31 Heersmink, « Varieties of Artifacts », p. 580.
 - 32 Dennis Muller, Maurice Chiodo et Mitja Siemknecht, « Integrators at War : Mediating in AI-assisted Resort-to-Force Decisions », 12 janvier 2025, p. 7. Source originale : https://www.researchgate.net/publication/387976034_Integrators_at_War_Mediating_in_AI-assisted_Resort-to-Force_Decisions, consulté le 7 octobre 2025.
 - 33 Heersmink, « Dimensions of Integration », p. 582.
 - 34 Karina Vold et José Hernández-Orallo, « AI Extenders : The Ethical and Societal Implications of Humans Cognitively Extended by AI », *AIES '19 Session 9 : Interaction homme-machine*, 2019, p. 507.
 - 35 Il convient de noter que le terme « transparence » tel qu'il est utilisé par Wheeler fait référence à un outil qui échappe à la conscience de l'utilisateur. Cela diffère de l'utilisation du terme « transparence » dans la littérature sur l'IA, où il fait référence à la capacité de comprendre le fonctionnement interne d'une IA.
 - 36 Heersmink, « Varieties of Artifacts », p. 577.
 - 37 Michael Wheeler, « The Reappearing Tool: Transparency, Smart Technology, and the Extended Mind », *AI & Society* 34 (2019), p. 857 à 866.
 - 38 Wheeler, « The Reappearing Tool », p. 857-866.
 - 39 Wheeler, « The Reappearing Tool », p. 864.
 - 40 Wheeler, « The Reappearing Tool », p. 864.
 - 41 H. Fowler, *Platon : Phaedrus*, Loeb Classical Library, 1971, 275^e.
 - 42 Comité international de la Croix-Rouge, « The Element of Human Control », *Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*, 2018, p. 2. Source originale : https://reachingcriticalwill.org/images/documents/Disalibr-fora/ccw/2018/hcp-meeting/documents/ICRC_WP3.pdf, consulté le 6 octobre 2025.
 - 43 É.-U., « Human-machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems », *Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*, 2018, p. 7. Source originale : <https://reachingcriticalwill.org/images/documents/Disalibr-fora/ccw/2018/gge/documents/GGE.2-WP4.pdf>, consulté le 6 octobre 2025.
 - 44 Australie, « Australia's System of Control and Applications for Autonomous Weapon Systems », *Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*, 2018, p. 5. Source originale : <https://reachingcriticalwill.org/images/documents/Disalibr-fora/ccw/2019/gge/Documents/GGE.2-WP2.pdf>, consulté le 6 octobre 2025.
 - 45 James Kraska, « Command Accountability for AI Weapon Systems in the Law of Armed Conflict », *International Law Study* 97 (2021), p. 408 à 445.
 - 46 Danièle Amoroso et Guglielmo Tamburrini, « Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues », *Current Robotics Report* 1 (2020), p. 189.
 - 47 Douglas MacGregor, *Warrior's Rage : The Great Tank Battle of 73 Easting*, Annapolis, Brookings Institution Press, 2009.
 - 48 Markus Langer, Cornelius J. König, Caroline Back et Caroline Hemming, « Trust in Artificial Intelligence : Comparing Trust Processes Between Human and Automated Trustees in Light of Unfair Bias », *Journal of Business and Psychology* 38 (2023), p. 493.
 - 49 Maakie Verbruggen, « No, Not That Verification: Challenges Posed by Testing, Evaluation, Validation and Verification of Artificial Intelligence in Weapon Systems », dans *Armament, Arms Control and Artificial Intelligence: The Janus-Faced Nature of Machine Learning in the Military Realm*, sous la dir. de Thomas Reinhold et de Niklas Schörning, Springer International Publishing AG, 2022, *ProQuest Ebook Central*, <http://ebookcentral.proquest.com/lib/cvlibrary-ebooks/detail.action?docId=7108384>. Créé à partir de cvlibrary-ebooks le 2025-07-09 18:36:43, p. 242.
 - 50 Commandant de l'Armée canadienne, *INSTRUCTION POUR LES OPÉRATIONS TERRESTRES* (2014), 6-3-11.
 - 51 Commandant de l'Armée canadienne, *INSTRUCTION POUR LES OPÉRATIONS TERRESTRES*, 6-3-12.
 - 52 Wheeler, « The Reappearing Tool », p. 859.
 - 53 Shan Carter et Michael Nielsen, « Using Artificial Intelligence to Augment Human Intelligence », *Distill*, 2017. Source originale : <https://distill.pub/2017/aia>, consulté le 3 novembre 2025.