L’IA et la transformation des guerres contemporaines

Défis du contrôle humain et du droit international humanitaire

Par Daniel Guinier

Expert de justice honoraire
Ancien expert devant la Cour pénale Internationale de La Haye
Chargé d'enseignement universitaire et conférencier

Et si la décision de tuer venait progressivement à se déplacer de l’humain vers le code ? Avec l’essor de l’intelligence artificielle (IA) et des systèmes d’armes autonomes, la guerre entre dans une nouvelle ère, où l’automatisation croissante des processus décisionnels militaires redéfinit les rapports entre l’humain, la machine et la responsabilité dans l’usage de la force. Les algorithmes reconfigurent la chaîne décisionnelle, brouillent les mécanismes d’imputabilité et mettent à l’épreuve le droit international humanitaire (DIH), entre automatisation de la violence et exigence d’un contrôle humain significatif.

Les guerres ont souvent constitué un puissant facteur d'accélération des innovations technologiques. Après les révolutions industrielle, nucléaire et numérique, l'introduction de l'IA marque une nouvelle étape, dans laquelle la machine ne se limite plus à assister l'action humaine, mais intervient directement dans le processus de décision militaire. Les systèmes d'armes létaux autonomes (LAWS)[1], capables d'identifier, de sélectionner et d'engager des cibles avec un degré variable d'autonomie, transfèrent progressivement une partie du jugement humain vers les algorithmes. Cette évolution soulève des interrogations majeures quant aux fondements de la responsabilité, de la chaîne de commandement et de la légitimité du recours à la force.

L'emploi des drones illustre cette mutation. Initialement dédiés à la reconnaissance, ils s'intègrent désormais à des chaînes décisionnelles où l'IA participe à la détection, au ciblage et à l'engagement des objectifs. L'expérimentation rapportée par Matthew Sparkes (2026), fondée sur le témoignage d'Alexander Kokhanovskyy, de drones entièrement autonomes capables de neutraliser une cible sans supervision humaine directe[2], constituerait, si elle était confirmée, un tournant majeur.

Les lois fondamentales de la robotique d'Isaac Asimov (1950)[3] ne sont pas transposables aux LAWS[4]. En effet, ces systèmes obéissent à des objectifs opérationnels et à des règles d'engagement plutôt qu'à des principes moraux

[1] Les LAWS, *(Lethal Autonomous Weapon Systems)*, regroupent des systèmes d'armes tels que les drones, les robots terrestres et navals, utilisant l’IA embarquée pour la surveillance, la détection et l’engagement de cibles.

[2] S'agissant d'une déclaration rapportée à l'issue d'une interview, et non d'une source institutionnelle ou scientifiquement vérifiée, cette affirmation doit, à ce stade, être appréhendée avec prudence. Il avait déjà été rapporté, dans le même magazine, le 13 octobre 2023, que des drones ukrainiens équipés d'IA, qualifiés de *"robots tueurs"*, capables d'identifier et d'attaquer des cibles sans aucun contrôle humain, auraient été utilisés sur un champ de bataille.

[3] 1ère loi : Un robot ne peut porter atteinte à un être humain ni, par son inaction, permettre qu’un humain soit blessé. 2ème loi : Un robot doit obéir aux ordres humains, sauf si ces ordres contredisent la première loi. 3ème loi : Un robot doit protéger sa propre existence tant que cela n’entre pas en conflit avec les deux premières lois.

[4] Nous proposons de les reformuler ainsi : 1ère loi : Ne pas engager une cible tant qu’elle n’a pas été classée comme hostile selon un modèle probabiliste et les règles d’engagement en vigueur. *L’interdiction morale absolue laisse ici place à un seuil statistique de ciblage.* 2ème loi : Exécuter la mission assignée dans le cadre des paramètres définis par la chaîne de commandement. *Si l’IA*

© 2026 Daniel Guinier

universels. La responsabilité devient alors distribuée entre concepteurs, opérateurs, autorités militaires et systèmes techniques. Le DIH demeure applicable, mais il a été conçu pour des décisions létales prises par des êtres humains. L'autonomisation des systèmes d'armes remet en question l'imputabilité des décisions ainsi que le maintien d'un contrôle humain effectif, alors même que les principes de distinction, de proportionnalité, de précaution et d'humanité reposent sur l'exercice d'un jugement humain.

Le présent article analyse les transformations induites par l'IA dans la décision létale et évalue dans quelle mesure l'évolution des systèmes d'armes autonomes révèle les limites du cadre juridique international, notamment en matière de contrôle humain, d'attribution des responsabilités et de cybersécurité.

L'IA et la recomposition de la décision létale

La guerre à distance, fondée sur l'emploi de missiles, de drones, de systèmes de surveillance et d'algorithmes d'aide à la décision transforme simultanément les modalités du combat, la décision létale et les responsabilités qui en découlent. L'intégration de l'IA automatise progressivement la détection, l'identification et la classification des cibles, accélérant les cycles décisionnels tout en renforçant la dépendance aux données et aux interfaces numériques. Si cette évolution améliore la réactivité opérationnelle, elle réduit le temps consacré à l'appréciation humaine et accroît les risques liés aux limites des modèles algorithmiques.

La décision létale ne relève plus d'un décideur unique mais d'une chaîne distribuée associant capteurs, systèmes embarqués, infrastructures de commandement, algorithmes d'IA et opérateurs humains. Trois fonctions principales sont à distinguer : **le guidage**, qui assure la navigation de l'arme ; **le ciblage**, chargé de détecter, identifier et sélectionner les objectifs ; et **la décision létale**, qui valide ou déclenche l'emploi de la force. Elles s'articulent désormais au sein d'architectures complexes qui diffusent les responsabilités entre de multiples acteurs. Andreas Matthias (2004) décrit cette situation comme une zone d'indétermination de la responsabilité, soulignant la nécessité d'un contrôle humain effectif.

Les relations entre l'humain et la machine s'organisent autour de **trois modèles** : l'humain valide chaque frappe *("human in the loop")*, supervise un système autonome *("human on the loop")*, ou est clairement absent de la décision *("human out of the loop")*. De plus, le passage entre ces modèles s'effectue progressivement, au rythme de l'autonomisation croissante des systèmes, au vu des pratiques opérationnelles qui illustrent cette évolution.

Concernant les **États-Unis**, les drones *MQ-9 Reaper* reposent sur une architecture où la fusion de données alimente le ciblage tandis que la décision létale demeure officiellement sous contrôle humain (US Department of Defense (2023)). Pour **Israël,** l'intégration du renseignement, de l'IA et du ciblage automatisé accélère la production des objectifs, réduisant le

identifie une opportunité de frappe plus efficace, elle peut la suggérer ou la prioriser, l'humain conservant son rôle de supervision.
3ème loi : Assurer sa propre continuité opérationnelle en se protégeant contre le brouillage, les cyberattaques et la destruction physique, à condition de ne pas compromettre la mission. *Dans ce cadre, l'IA militaire privilégie la continuité de la mission plutôt que sa propre survie.*

© 2026 Daniel Guinier

temps disponible pour leur validation (ICRC (2016)). Le conflit en **Ukraine** illustre une transition rapide vers des systèmes semi-autonomes. Si les drones demeurent pilotés par des opérateurs (Jack Watling *et al.* (2023)), l'intégration de la vision artificielle et du verrouillage automatique transfère progressivement la décision à la machine. Les expérimentations de systèmes entièrement autonomes pourraient ainsi marquer le passage vers une décision létale sans intervention humaine directe.

Cette évolution complexifie l'application du DIH. La multiplication des acteurs : *concepteurs, développeurs, analystes, opérateurs et chaîne de commandement,* rend plus difficile l'attribution des responsabilités (CICR (2020)). Les conséquences ne sont pas uniquement juridiques. Malgré l'éloignement du champ de bataille, les opérateurs demeurent exposés à une blessure morale liée à leur participation à l'usage de la force (Brett Litz *et al.* (2009)).

Enfin, la convergence entre IA, drones, cyberdéfense et réseaux numériques favorise une guerre en réseau où la décision est distribuée entre acteurs humains et systèmes autonomes. Les liens croissants entre technologies civiles et militaires, notamment à travers les environnements de simulation et les jeux vidéo, participent également à cette transformation (Markus Schulzke (2013)).

Le droit international humanitaire face à l'autonomisation de la décision létale

L'intégration de l'IA dans les systèmes d'armes met en évidence les limites du droit international humanitaire (DIH). Si ce dernier reste applicable à l'ensemble des moyens et méthodes de guerre, il a été conçu dans un contexte où la décision de recourir à la force relevait exclusivement d'un acteur humain. L'automatisation de la détection, de l'identification, de la sélection et parfois de l'engagement des cibles fragilise les mécanismes traditionnels d'imputabilité, la décision létale résultant désormais d'une chaîne distribuée associant capteurs, algorithmes, opérateurs et autorités militaires (CICR (2020)).

Cette difficulté est renforcée par les caractéristiques des systèmes fondés sur l'apprentissage automatique. Leur fonctionnement, parfois opaque, complique les exigences de traçabilité, de justification et de contrôle indispensables à l'application des principes de distinction, de proportionnalité et de précaution. L'accélération des cycles décisionnels réduit également le temps disponible pour l'analyse juridique et le jugement humain, pourtant au cœur de ces principes (Vincent Boulanin *et al.* (2021)).

Le cadre juridique actuel repose principalement sur les Conventions de Genève, leurs Protocoles additionnels et l'Art. 36 du Protocole additionnel I (1977), qui impose une évaluation préalable des nouvelles armes. Toutefois, cette procédure apparaît insuffisamment adaptée aux systèmes d'IA dont les comportements peuvent évoluer au cours de leur emploi.

Les négociations conduites dans le cadre de la Convention sur certaines armes classiques (CCAC) illustrent les divergences persistantes entre les États quant à une interdiction ou à un encadrement des systèmes d'armes autonomes. Human Rights Watch (2020) considère que la délégation d'une décision létale à une machine est incompatible avec les exigences fondamentales des droits humains. Le CICR préconise pour sa part l'interdiction des systèmes imprévisibles et un encadrement strict des systèmes autonomes visant des personnes (CICR (2021)). Toutefois, la notion désormais centrale de **contrôle humain effectif** reste dépourvue de définition juridiquement contraignante.

Les conflits récents, notamment en Ukraine, illustrent le décalage croissant entre la rapidité des évolutions technologiques et celle du droit. L'enjeu ne réside plus dans la création de nouveaux principes, mais dans

© 2026 Daniel Guinier

l'adaptation de leur mise en œuvre afin de garantir un contrôle humain effectif, une responsabilité identifiable et une transparence suffisante des processus décisionnels.

La cybersécurité comme autre condition du contrôle humain effectif

La cybersécurité constitue désormais une composante essentielle de l'emploi des systèmes d'armes intégrant l'IA. La problématique s'inscrit pleinement dans la mutation des risques associés aux systèmes multi-agents autonomes d'IA, telle qu'analysée par Daniel Guinier (2026). Ces derniers ne résultent plus uniquement des limites des algorithmes, mais aussi des vulnérabilités affectant les données, les capteurs, les communications, ou les architectures numériques. Une compromission peut conduire à une mauvaise identification des objectifs, à une altération des priorités opérationnelles ou à un emploi de la force contraire aux règles d'engagement.

La fiabilité de la décision létale dépend ainsi autant de l'intégrité des données et de la résilience des systèmes que des performances des algorithmes. Les attaques par manipulation des données (*data poisoning*), usurpation de signaux (*spoofing*), brouillage (*jamming*) ou intrusion dans les systèmes compromettent directement le maintien d'un contrôle humain effectif ainsi que le respect des principes du DIH (CICR (2020) ; NATO CCDCOE (2026)).

Ces vulnérabilités prolongent les difficultés déjà soulevées par la fragmentation de la chaîne décisionnelle. Une attaque cyber peut interrompre l'exercice du jugement humain, modifier le comportement d'un système autonome et rendre particulièrement complexe l'identification du responsable d'une violation du DIH. Les insuffisances du cadre juridique ne tiennent donc pas seulement à l'autonomie croissante des systèmes d'armes, mais aussi à leur dépendance à des infrastructures numériques vulnérables.

Conclusion et perspective

L'intégration croissante de l'IA et des systèmes algorithmiques dans les opérations militaires ne constitue pas une simple évolution technologique : elle transforme profondément les modalités de conduite de la guerre. En intervenant dans la collecte et l'analyse du renseignement, l'identification des cibles, l'aide à la décision et la priorisation des menaces, l'IA modifie le rapport entre l'homme, la machine et l'environnement conflictuel. Le champ de bataille devient ainsi un espace fortement structuré par les données, où la perception et l'action militaires reposent de plus en plus sur des systèmes capables de traiter des volumes d'informations dépassant les capacités humaines.

Cette transformation améliore la rapidité et l'efficacité opérationnelle, mais elle introduit également de nouvelles vulnérabilités liées à la dépendance technologique, aux biais algorithmiques et au risque de perte de contrôle dans les situations critiques. La question centrale n'est donc plus uniquement celle de la performance des systèmes d'armes, mais celle du degré d'autonomie acceptable dans la décision d'employer la force et du maintien d'une responsabilité humaine effective.

Le cadre juridique international apparaît aujourd'hui insuffisant pour répondre pleinement à ces évolutions. Bien que les principes du DIH : *distinction, proportionnalité et précaution*, demeurent applicables, leur mise en œuvre devient plus complexe lorsque les décisions militaires résultent d'une interaction entre opérateurs, capteurs, logiciels et infrastructures numériques. L'absence de règles précises concernant l'autonomie des systèmes d'armes et l'attribution des responsabilités en cas d'erreur algorithmique constitue ainsi un défi majeur.

© 2026 Daniel Guinier

Face à ces enjeux, l'adaptation du cadre normatif international apparaît nécessaire afin d'encadrer l'usage militaire de l'IA tout en préservant l'innovation technologique. Cette régulation devra s'accompagner d'une gouvernance fondée sur la transparence, la traçabilité et la possibilité de contrôler ou d'interrompre les processus automatisés. La capacité à comprendre les limites et les conditions de fiabilité des systèmes algorithmiques devient une exigence essentielle de légitimité opérationnelle.

Enfin, le recours croissant au combat augmenté par l'IA invite à repenser la nature même de la guerre contemporaine. La convergence entre simulation, jeux vidéo, cybersécurité et technologies militaires montre que l'information, le calcul et la perception constituent désormais des ressources stratégiques majeures. Le défi ne consiste donc pas seulement à développer des machines plus performantes, mais à définir les conditions permettant leur intégration dans un cadre stratégique, juridique et éthique garantissant le maintien du contrôle humain et de la responsabilité politique dans l'usage de la force.

Avec l'IA, **le soldat pourrait moins viser que superviser**. Il ne sera plus seulement celui qui appuie sur la gâchette, mais **un informaticien** chargé d'arbitrer et d'interrompre, si nécessaire, l'engagement proposé par un algorithme… à condition qu'il conserve une capacité réelle de contrôle et ne se réduise pas à un simple validateur humain d'une décision construite par la machine.

Bibliographie sélective

Asimov, Isaac (1950) : I, Robot, Gnome Press, New York ; Réédition de 2018, HarperVoyager, 256 p.

Boulanin, Vincent *et al.* (2021) : Autonomous Weapon Systems and International Humanitarian Law : Identifying Limits and the Required Type and Degree of Human-Machine Interaction, SIPRI, Stockholm, 66 p.

CICR (2020) : Le droit international humanitaire et les défis des conflits contemporains, Comité international de la Croix-Rouge, Genève, 120 p.

CICR (2021) : Position du CICR sur les systèmes d'armes autonomes, Genève, 13 p.

Human Rights Watch (2020) : Stopping Killer Robots. Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control. Rapport, août, 62 p.

Guinier, Daniel (2026) : Réseaux d'agents d'IA autonomes - Mutation des risques et du pouvoir à l'ère agentique. Expertises, n°525, juillet, pp. 23-26.

Litz, Brett T. *et al.* (2009) : Moral injury and moral repair in war veterans: a preliminary model and intervention strategy, Clinical Psychology Review, 29, pp. 695–706.

Matthias, Andreas (2004) : The Responsibility Gap : Ascribing Responsibility for Actions of Learning Automata. Ethics and Information Technology, pp. 175-183.

NATO CCDCOE (2026) : Locked Shields Exercise Reports, Cooperative Cyber Defence Centre of Excellence, Tallinn.

Schulzke, Marcus (2013) : Simulating War : Studying Conflict Through Simulation Games, Journal of Military Ethics, vol. 12, n°3, Routledge, pp. 188–199.

Sparkes, Matthew (2026) : Fully autonomous drones have killed human soldiers for the first time. New Scientist, 10 juin.

US Department of Defense (2023) : Directive 3000.09 : Autonomy in Weapon Systems, DoD, 24 p.

Watling, Jack *et al.* (2023) : Preliminary Lessons from Russia's Unconventional Operations During the Russo-Ukrainian War, February 2022–2023. RUSI Special Report, 29 mars, 36 p.

© 2026 Daniel Guinier