

Terminator Ethics : faut-il interdire les « robots tueurs » ?

Par **Jean-Baptiste Jeangène Vilmer**

Jean-Baptiste Jeangène Vilmer, docteur en science politique et en philosophie, juriste, enseigne les relations internationales, l'éthique et le droit de la guerre à Sciences Po Paris. Membre de l'Institut international de droit humanitaire de San Remo, il est l'auteur d'une quinzaine de livres, dont *La Guerre au nom de l'humanité* (Paris, PUF, 2012), distingué par le prix Maréchal Foch de l'Académie française¹.

Les systèmes d'armes létaux autonomes font débat, certaines ONG réclamant leur interdiction préventive. La notion d'autonomie est pourtant très complexe : il n'existe pas de système d'arme totalement soustrait à l'intervention humaine. En termes moraux et vis-à-vis du droit international humanitaire, ces systèmes posent des problèmes peu différents des autres armes utilisées par l'homme. Il semble donc préférable de les encadrer, plutôt que les interdire *a priori*.

politique étrangère

Les systèmes d'armes létaux autonomes (SALA), souvent appelés « robots tueurs », sont théoriquement dotés d'une capacité de désigner leur cible et d'ouvrir le feu sans intervention ni supervision humaine. Leur développement s'inscrit dans une tendance générale et inévitable à la robotisation militaire². Leurs avantages potentiels sont économiques – réduction des coûts et du personnel –, opérationnels – ils accroissent la rapidité de la prise de décision et réduisent la dépendance à l'égard des communications –, sécuritaires – remplaçant ou assistant les humains, ils minimisent pour eux les risques –, voire humanitaires – ils pourraient à terme mieux respecter le droit de la guerre que les humains.

Ils suscitent déjà un vaste débat, lancé par les organisations non gouvernementales (ONG)³ et le rapporteur spécial sur les exécutions extrajudiciaires,

1. Les vues exprimées dans cet article sont celles de l'auteur et ne sauraient engager une quelconque institution.

2. J.-B. Jeangène Vilmer, « Robotisation et transformations de la guerre », *Politique étrangère*, vol. 78, n° 3, automne 2013, p. 80-89.

3. Depuis 2009, l'International Committee for Robots Arms Control (ICRAC), puis Human Rights Watch (HRW), Article 36 et PAX, qui font désormais partie, avec une dizaine d'autres ONG, de la *Campaign to Stop Killer Robots* lancée en 2013. Voir sur : <www.stopkillerrobots.org>.

sommaires ou arbitraires de l'Organisation des Nations unies (ONU)⁴, qui demandent leur interdiction préventive. La France a organisé et présidé à l'ONU en mai 2014 la première réunion multilatérale sur le sujet, où se sont exprimées 24 délégations nationales et plusieurs ONG⁵ ; d'autres rencontres internationales ont eu lieu depuis.

De quoi parle-t-on ?

Le débat est rendu confus par trois difficultés originelles. Une ambiguïté sémantique d'abord : les termes employés sont très divers (robots tueurs, robots létaux autonomes, systèmes d'armes robotisés, systèmes d'armes autonomes, systèmes d'armes létaux autonomes), chacun y allant de sa définition, ce qui n'aide pas à la compréhension.

L'appellation « systèmes d'armes létaux autonomes », qui tend aujourd'hui à s'imposer, présente l'avantage – par rapport à « robots létaux autonomes » (RLA) – de ne pas reposer sur le terme ambigu de « robot ». Apparu en 1920 sous la plume de l'écrivain tchèque Karel Tchépek pour désigner des machines pouvant accomplir des tâches à la place de l'homme – le tchèque *robot* qualifiant un travail pénible, une corvée –, il n'en existe pas de définition consensuelle. Au sens large, un robot est une machine programmable dotée de capteurs et réagissant à son environnement.

Dans le cas du pilote automatique d'un aéronef ou d'un drone, l'homme reste présent : on parle d'un système robotisé. D'où l'appellation de « systèmes militaires robotisés » (SMR), également populaire, mais qui ne désigne pas la même chose. Les SALA ne sont pas des SMR : l'hypothèse sur laquelle ils reposent est précisément que l'homme n'est pas dans la boucle. Ce qui les distingue est donc leur degré d'autonomie.

Par ailleurs, l'adjectif « létaux » est facultatif. Le Comité international de la Croix-Rouge (CICR) estime que la létalité d'une arme dépend du contexte. D'autres soulignent que des adjectifs comme « létaux » ou « tueur » réduisent le débat au fait de tuer les humains, alors qu'il concerne aussi des systèmes pouvant seulement les blesser, ou causer des dommages matériels. On peut d'ailleurs se demander si le fait de remplacer les balles réelles par des munitions non létales, en caoutchouc par exemple – hypothèse réaliste car des systèmes plus ou moins autonomes pourraient être utilisés par la police antiémeute –, changerait le problème pour ceux qui s'opposent aux SALA par principe, et donc indépendamment des conséquences de leur usage.

4. En 2010, Philip Alston avait attiré l'attention sur le sujet et son successeur, Christofer Heyns, a demandé un moratoire en 2013.

5. Réunion informelle d'experts sur les systèmes d'armes létaux autonomes, Convention sur certaines armes classiques (CCAC), Organisation des Nations unies (ONU), Genève, 13-16 mai 2014.

La deuxième difficulté du débat tient à sa nature spéculative : il s'agit d'évaluer la légalité et la légitimité d'armes n'existant pas, et dont on ne peut prévoir l'évolution technologique. Ce qui nuit à la discussion, tout d'abord en laissant flotter une ambiguïté entre présent et futur, entre ce qui existe et ce qui est envisagé, voire fantasmé ; ensuite, en laissant libre cours à l'imagination de chacun, les défenseurs des SALA prévoyant que ceux-ci respecteront – même mieux que les humains – le droit international humanitaire (DIH), et leurs opposants que ce ne sera pas le cas, sans que personne ne puisse rien prouver ; enfin, en incitant les opposants à profiter de l'ambiguïté des termes pour intégrer au débat des armes existantes auxquelles ils s'opposent par ailleurs.

Si les ONG affectionnent l'appellation « robots tueurs » (*killer robots*) à destination du grand public⁶ – quand, dans les enceintes internationales, elles parlent le langage plus neutre des États (RLA ou SALA) –, ce n'est pas seulement parce qu'il s'agit d'un terme sensationnel, mais surtout parce que son acception est plus large : elle ne se limite pas aux machines autonomes et peut donc inclure des armes existantes comme des mines, des missiles et surtout des drones armés. Ainsi la confusion est-elle entretenue dans l'opinion entre les robots tueurs des films de science-fiction, qui servent de repoussoir, et les drones armés existants, qui sont pourtant téléopérés. Ce débordement du débat sur les drones, victimes du syndrome *Terminator*, relève de la désinformation et doit être combattu.

Troisième difficulté, la notion d'autonomie. En évacuant le terme de « robot », qui a le défaut de renvoyer dans l'imaginaire collectif aux humanoïdes propres à la science-fiction, les SALA n'ont pas résolu le problème de définition : la notion d'autonomie sur laquelle ils reposent n'est pas plus claire. Il convient d'y revenir plus longuement.

La notion d'autonomie

Il faut d'abord distinguer les systèmes automatiques, préprogrammés pour accomplir des tâches spécifiques dans un environnement prédéfini et contrôlé, des systèmes autonomes, qui décident si et quand accomplir cette tâche dans un environnement changeant et imprévisible⁷. Basés sur des règles, les premiers sont déterministes, donc prévisibles. Le

6. Elles ne sont toutefois pas les seules à utiliser cette expression, d'abord introduite par le philosophe Robert Sparrow (« Killer Robots », *Journal of Applied Philosophy*, vol. 24, n° 1, 2007, p. 62-77) et le politologue Armin Krishnan (*Killer Robots: Legality and Ethicality of Autonomous Weapons*, Farnham, Ashgate, 2009).

7. UK Ministry of Defence, Development, Concept and Doctrine Centre, *Joint Doctrine Publication 0-01.1: UK Supplement to the NATO Terminology Database*, septembre 2011, p. A2.

système *Phalanx* qui équipe l'US Navy depuis 1980, ou sa version terrestre *Counter Rocket, Artillery and Mortar (C-RAM)*, l'*Iron Dome* israélien et les armes amorcées par capteurs (*sensor-fused weapons*) sont plus perfectionnés qu'une machine distributrice de boissons mais fonctionnent sur le même modèle : ils font telle action à tel signal, inéluctablement et régulièrement. Les systèmes autonomes sont plus indépendants, en ce qu'ils jouissent d'une certaine liberté de comportement, et du même coup moins prévisibles.

Ces deux catégories ne sont ni homogènes – il y a des niveaux d'automatisme et des niveaux d'autonomie –, ni exclusives l'une de l'autre. Il n'y a pas de frontière indiscutable entre l'automatisme et l'autonomie, plutôt un continuum ; et les systèmes envisagés peuvent être multimodes, donc hybrides. Ils peuvent être automatiques pour certaines fonctions (ciblage, tir) et autonomes pour d'autres (déplacement). La question n'est pas ici celle de la navigation, mais du ciblage et de l'ouverture du feu.

Les armes plus ou moins autonomes sont souvent divisées entre 1) armes semi-autonomes (*human-in-the-loop weapons*), pour lesquelles la décision d'ouverture du feu reste humaine, et dont on considère acceptable l'usage légal non seulement défensif mais aussi offensif (munitions à tête chercheuse, drones armés, missiles balistiques intercontinentaux, etc.) ; 2) armes autonomes supervisées (*human-on-the-loop weapons*), qui désignent et traitent seules leurs cibles, tout en restant sous la supervision d'un humain pouvant les arrêter, et dont on considère acceptable l'usage légal défensif seulement, et pour cibler des matériels et non directement des humains (systèmes de défense antimissile) ; 3) armes autonomes (*human-out-of-the-loop weapons*), qui désignent et traitent seules leurs cibles, ne sont pas supervisées, et dont seul un usage non légal contre des cibles matérielles est considéré acceptable (systèmes de brouillage électronique). La question que pose l'hypothèse des SALA est précisément de savoir si l'on peut faire un usage légal de cette autonomie.

La place de l'homme dans la boucle

Cette présentation commune simplifie les choses et ne rend pas compte du fait que l'autonomie n'est pas un escalier à trois marches, mais un continuum de multiples degrés. Les passages du contrôle « total » au contrôle « partiel » puis à l'absence de contrôle sont flous, et les armes évoluent. Pour autant, la division *in/on/out of the loop* permet de comprendre que les SALA ne sont pas des machines toujours et par défaut autonomes, mais des systèmes multimodes pouvant passer d'un mode téléopéré à un mode sous supervision, puis sans supervision – les deux derniers déléguant le tir, le dernier (mode SALA) le faisant sans contrôle humain.

La définition

D'aucuns préfèrent renoncer à toute tentative de définition, croyant ainsi éviter le débat aujourd'hui pour protéger leurs intérêts demain. Mauvais calcul. Il faut au contraire définir, et restrictivement, pour éviter l'amalgame avec des technologies existantes (drones armés, missiles ou systèmes de défense antimissiles).

On peut définir le SALA comme *un système d'arme qui, une fois activé, est capable de décider seul, c'est-à-dire sans intervention ni supervision humaine, du ciblage et du déclenchement de la frappe, en fonction d'un environnement changeant auquel il s'adapte*. Une version plus synthétique est utilisée par le gouvernement américain comme par le rapporteur spécial de l'ONU ou par Human Rights Watch (HRW) : « un système d'arme qui, une fois activé, permet de sélectionner et de traiter des cibles sans intervention d'un opérateur humain⁸. »

Le faux problème du système d'arme « pleinement » autonome

La précision « une fois activé » est importante car, avant l'activation, il y a nécessairement intervention humaine, dans la conception, la programmation et le déploiement. La pertinence de l'expression « pleinement autonome » (*fully autonomous weapons*) est donc relative puisque tout système actuel ou futur dépendra d'une intervention humaine initiale. À moins que la programmation elle-même s'automatise et qu'à terme les machines soient conçues par d'autres machines – un scénario de film de science-fiction qui pourrait seul justifier l'adverbe « pleinement » autonome.

Parler d'une arme pleinement autonome est donc un abus de langage, que pratiquent certains fabricants et États pour vanter leurs matériels. Ainsi le ministère britannique de la Défense a-t-il présenté le prototype de drone de combat *Taranis* comme tel, sous-entendant qu'il pourrait choisir ses cibles et frapper sans intervention ni supervision humaine – ce qui n'est pas établi. BAE Systems et la Royal Air Force présentent également le *Brimstone*⁹ de MBDA comme pleinement autonome alors que, comme pour les armes amorcées par capteurs, il y a bien un contrôle humain dans le paramétrage du ciblage et le choix de la zone. La fameuse sentinelle immobile *Samsung SGR-A1* déployée à la frontière des deux Corées est aussi réputée autonome alors qu'elle a plutôt une commande automatique

8. UN Doc A/HRC/23/47, 9 avril 2013, § 38 (le texte original est au pluriel). Voir US Department of Defense, « Autonomy in Weapons Systems », directive n° 3000.09, 21 novembre 2012, glossaire, deuxième partie ; UK Ministry of Defence, « The UK Approach to Unmanned Aircraft Systems », par. 202 et 203 ; HRW, *Losing Humanity: The Case Against Killer Robots*, 2012, p. 2.

9. Il s'agit d'un missile air-sol de type *fire and forget*, autonome après lancement.

débrayable d'ouverture du feu. Le ministère russe de la Défense a quant à lui présenté son « complexe robotique mobile », développé par Izhevsk Radio Plant et qui garde les installations de missiles balistiques, comme capable de cibler et tirer sans intervention ni supervision humaine, ce qui n'est pas non plus établi.

Les deux seuls pays ayant une politique officielle en la matière, les États-Unis et le Royaume-Uni, reconnaissent ne pas vouloir de systèmes pleinement autonomes : pour les premiers, « les systèmes d'armes autonomes et semi-autonomes doivent être conçus de façon à permettre aux commandants et aux opérateurs d'exercer des niveaux appropriés de jugement humain sur l'usage de la force¹⁰ » ; pour le second, « les systèmes d'armes seront toujours sous contrôle humain¹¹ ».

Les Américains ne précisent pas ce qu'est un niveau « approprié » de jugement humain et les Britanniques considèrent que le contrôle humain peut se manifester dans la programmation – ce qui signifie qu'un système qui agit sans supervision humaine peut être considéré « sous contrôle humain » dans la mesure où il a été programmé par l'homme.

Produire des armes « totalement » autonomes n'est dans l'intérêt de personne, mais c'est pourtant avec cela que l'on veut effrayer le grand public. Un sondage conduit par Charli Carpenter, de l'université du Massachusetts, conclut qu'une majorité des sondés américains sont « fermement opposés » à l'usage de robots létaux totalement autonomes

Une simplification du débat

– les plus opposés étant d'ailleurs les militaires¹². Citant cette étude pour illustrer la « réaction d'incrédulité et d'horreur, presque toujours suivie d'un choc sérieux » qui frappe le public lorsqu'il découvre ce problème, l'ONG PAX supprime au passage le « totalement » et parle de l'opposition des sondés aux robots autonomes¹³. Les opposants veulent profiter de l'opposition consensuelle à la pleine autonomie – qui n'existe pas –, réunissant ONG, gouvernements, civils et militaires, pour faire l'amalgame avec toute forme d'autonomie et inclure dans le débat des armes existantes.

10. US Department of Defense, directive citée, § 4(a).

11. UK Ministry of Defence, *Written Evidence from the Ministry of Defence submitted to the House of Commons Defence Committee inquiry "Remote Control: Remotely Piloted Air Systems – current and future UK use"*, septembre 2013, p. 3.

12. C. Carpenter, « US Public Opinion on Autonomous Weapons », 2013, disponible sur : <www.whiteo-liphaunt.com/duckofminerva/wp-content/uploads/2013/06/UMass-Survey_Public-Opinion-on-Autonomous-Weapons.pdf>.

13. PAX, *Deadly Decisions: 8 Objections to Killer Robots*, 2014, p. 6.

À l'inverse, l'une des stratégies des États pour évacuer le problème est de définir les SALA de façon très restrictive comme systèmes d'armes létaux pleinement autonomes, pour ensuite reconnaître que de telles armes incontrôlables ont peu d'intérêt opérationnel et ne seront donc sans doute jamais construites¹⁴. Les ONG avaient vu venir cette stratégie : « un des principaux obstacles auxquels la Campagne [pour interdire les robots tueurs] devra faire face sera sans doute que des États soutiendront l'idée d'une interdiction mais mettront le seuil de l'autonomie si haut que cela n'affectera aucun des systèmes robotiques qu'ils souhaitent déployer¹⁵. »

Les ONG ont compris que la question n'est pas la légitimité ou la légalité d'armes pleinement autonomes qui ne seront jamais produites, mais celle d'armes relativement autonomes, et du degré de contrôle humain nécessaire. À l'ONU, les États n'ont cessé de dire qu'il fallait « un homme dans la boucle », un « contrôle humain » ; et les ONG n'ont cessé de leur demander ce que cela signifiait exactement. Certains États, comme l'Irlande, ont précisé que ce contrôle humain ne devait pas être symbolique mais suffisant, adéquat et significatif.

Les ONG ont donc délaissé le combat contre les armes pleinement autonomes pour se concentrer sur la notion de « contrôle humain significatif » (*meaningful human control*). L'exemple généralement donné d'un contrôle humain insignifiant est celui d'un opérateur qui appuie sur le bouton « feu » chaque fois qu'un signal lumineux s'affiche, sans aucune autre information ; pour exercer un contrôle humain significatif, il devrait tenir compte d'informations concernant la cible, le contexte et les effets probables de la frappe.

Les États voient cette évolution d'un mauvais œil, puisqu'elle concerne un éventail large de technologies, comprenant des systèmes existants qui dépendent tous d'un certain degré de contrôle humain. Les appels à éviter ce débat sont vains – il est déjà ouvert –, et la crainte qui les motive n'est pas justifiée. Comme tous les concepts de la théorie de la guerre juste (autorité légitime, juste cause, intention droite, dernier recours, effet positif, proportionnalité), la notion de contrôle humain « significatif » est en soi vague et discutable. Elle donnera lieu à des interprétations divergentes, sans aucune chance de susciter un consensus pouvant fonder l'interdiction préventive que réclament les ONG.

14. Les Japonais, par exemple, considèrent les SALA comme « des systèmes d'armes létaux “pleinement” autonomes qui, une fois activés, peuvent choisir et attaquer une cible sans intervention humaine » et ils « ne sont pas convaincus du besoin de développer des systèmes d'armes létaux “pleinement” autonomes qui sont complètement hors de contrôle humain » (déclaration de l'ambassadeur Toshio Sano à la réunion de la CCAC citée, Genève, 13 mai 2014).

15. N. Marsh, *Defining the Scope of Autonomy*, PRIO Policy Brief, février 2014, p. 3.

Le débat moral

De nombreuses ONG et certains États comme le Pakistan ou Cuba demandent une interdiction préventive des SALA. Ils sont motivés par des raisons déontologistes et/ou conséquentialistes¹⁶. Du côté déontologiste, certains philosophes tels Peter Asaro ou Robert Sparrow, la plupart des ONG et le Vatican estiment que déléguer le choix du ciblage et de l'ouverture du feu à une machine « enfreint » la dignité humaine et que les hommes ont le « droit de ne pas être tués par une machine ». Ils citent abondamment la clause de Martens¹⁷.

Le rapporteur spécial de l'ONU exprime également une approche déontologiste lorsqu'il écrit que, quand bien même il pourrait « être prouvé que dans l'ensemble, ces robots pourront sauver des vies, la question est de savoir s'il n'est pas *intrinsèquement mauvais* de laisser des machines décider elles-mêmes qui et quand tuer. [...] La question est ici de savoir si l'emploi de RLA pour combattre des personnes, notamment des combattants ennemis, est acceptable *sur le principe*, parce qu'il s'agit de confier à des machines la décision de recourir à la force létale. Si la réponse est négative, aucun autre argument ne peut justifier l'emploi de RLA, quel que soit leur degré de compétence technique¹⁸. »

Ceux qui répondent clairement qu'il est par principe mauvais et inacceptable de déléguer l'ouverture du feu à une machine, le posent comme une pétition de principe. Ils ne définissent pas la dignité humaine invoquée, ni n'expliquent en quoi exactement elle serait violée. Quant à la clause de Martens, elle est plus un rappel – dans l'hypothèse où ces technologies ne seraient couvertes par aucune convention particulière, elles seraient tout de même sujettes à d'autres normes internationales – qu'un critère à prendre au pied de la lettre, et elle ne suffit certainement pas à justifier une interdiction.

Si la cible est légale et légitime, la question de savoir qui la tue (un humain ou une machine) a-t-elle une pertinence morale ? Et est-ce bien la

16. Le déontologisme est une approche d'origine kantienne selon laquelle une action est moralement bonne si elle est accomplie par devoir, par respect pour une norme, un principe. Elle suppose l'existence objective de certaines obligations morales universelles, et considère que les actes ont une valeur intrinsèque, indépendamment des conséquences. Le conséquentialisme est une approche d'origine anglaise (Jeremy Bentham, John Stuart Mill, Henry Sidgwick) selon laquelle une action est bonne lorsqu'elle produit les meilleures conséquences pour l'ensemble des individus concernés.

17. Préambule de la convention II de La Haye de 1899. Du nom du délégué russe ayant fait cette déclaration à la conférence, elle stipule : « En attendant qu'un code plus complet des lois de la guerre puisse être édicté, les Hautes Parties contractantes jugent opportun de constater que, dans les cas non compris dans les dispositions réglementaires adoptées par elles, les populations et les belligérants restent sous la sauvegarde et sous l'empire des principes du droit des gens, tels qu'ils résultent des usages établis entre nations civilisées, des lois de l'humanité et des exigences de la conscience publique. »

18. UN Doc. A/HRC/23/47, 9 avril 2013, § 92-93, nous soulignons.

machine qui la tue, ou l'humain qui l'a programmée ? Son autonomie n'est pas une autonomie de la volonté, une capacité de se donner à soi-même ses propres règles, mais plutôt une autonomie fonctionnelle, qui implique seulement de maîtriser les processus élémentaires (physiques et mentaux) permettant de réaliser la finalité poursuivie.

En outre, postuler comme le font les opposants déontologistes aux SALA qu'il est toujours pire d'être tué par une machine que par un homme, sans tenir compte des conséquences, peut conduire à des absurdités. L'approche déontologiste de Sparrow le contraint ainsi à conclure que les bombardements d'Hiroshima et de Nagasaki – qu'il ne justifie pas – sont plus humains car respectueux de la « dignité humaine » de leurs victimes que n'importe quel tir de SALA, pour la simple raison que les bombardiers étaient pilotés¹⁹...

Plus raisonnable est l'objection qui s'appuie non sur des principes obscurs mais sur les conséquences probables. Nombre de chercheurs, d'ONG mais aussi d'États pensent que les SALA en produiront de mauvaises, sur la communauté affectée et, *in fine*, sur l'efficacité de l'opération ; qu'ils seront donc contreproductifs pour la puissance utilisatrice.

D'abord, ils seraient incapables de « gagner les cœurs et les esprits ». On observe effectivement, par exemple avec les frappes de drones au Pakistan, qu'être tué *apparemment* par une machine – même si elle est téléopérée, et que donc c'est un homme qui appuie sur la gâchette – suscite une indignation et un sentiment d'injustice particuliers. Être tué par un missile tiré d'un avion piloté, d'un hélicoptère ou d'un croiseur à 1 500 kilomètres n'est pourtant pas différent, mais le drone suscite une réaction épidermique et largement irrationnelle dans l'opinion²⁰.

Les robots vont-ils remplacer les hommes ?

De plus, les SALA abaisseraient le seuil d'entrée en conflit : réduisant les coûts humains, ils inciteraient à faire la guerre. Les ONG imaginent déjà les gouvernements contourner le processus démocratique (vote au parlement) pour envoyer une armée de robots se battre à la place des hommes²¹.

19. Réponse de R. Sparrow à ma question lors de la conférence de l'International Studies Association (ISA), Toronto, 26-29 mars 2014.

20. J.-B. Jeangène Vilmer, « Légalité et légitimité des drones armés », *Politique étrangère*, vol. 78, n° 3, automne 2013, p. 119-132 et « Idéologie du drone », *La Vie des idées*, 4 décembre 2013, disponible sur : <www.laviedesidees.fr/IMG/pdf/20131204_jbjv_-_ideologie_du_drone-2.pdf>.

21. Par exemple PAX, *op. cit.*, p. 9.

Cette crainte partage avec la précédente une présomption de déshumanisation, la conviction que les robots sont là pour, à terme, remplacer les hommes. C'est ce que révèlent aussi des formules comme : « D'abord, il y avait des êtres humains sans machines. Ensuite, des êtres humains avec des machines. Et finalement, des machines sans êtres humains », que les ONG mettent en exergue²². Mais la robotisation militaire – qui d'ailleurs n'implique pas une baisse des personnels mais leur réaffectation (moins sur le terrain, plus à la programmation) – consiste à intégrer des robots dans des unités humaines : il ne s'agit pas de remplacer les humains dans toutes leurs missions, mais de les assister dans certaines d'entre elles.

Par ailleurs, ces SALA dilueraient la responsabilité. Elle sera certes plus compliquée, mais pas impossible à établir : la responsabilité étatique – de l'État ayant décidé de déployer l'arme – pourrait être engagée, mais aussi la responsabilité pénale individuelle du concepteur, du programmeur, du fabricant et du commandant – pour autant que la doctrine de la responsabilité du supérieur hiérarchique puisse, à certaines conditions, s'appliquer.

Enfin, la plus importante objection conséquentialiste est que ces armes ne pourront jamais respecter le droit international humanitaire (DIH), ce que pensent les ONG ainsi que nombre de chercheurs et plusieurs États (Pakistan, Autriche, Égypte et Mexique).

Le respect du droit international humanitaire

Il ne fait « aucun doute que le développement et l'usage de systèmes d'armes autonomes dans des conflits armés sont gouvernés par le DIH », estime le CICR²³. Les États le reconnaissent : ceux qui ont participé à la première réunion d'experts organisée à l'ONU en mai 2014 ont fait du respect du DIH une condition *sine qua non* de la mise en œuvre des SALA – avec des prédictions diverses, certains estimant que ceux-ci ne pourront passer ce test, d'autres soulignant la difficulté de se prononcer à ce stade sans connaître les capacités futures de ces armes (Japon, Australie). Tous ont également insisté sur la vérification *a priori*, avant leur mise en service, de leur conformité au DIH en vertu de l'article 36 du Protocole additionnel I aux conventions de Genève (PI)²⁴.

22. John Pike (GlobalSecurity.org), cité par F. Reed, « Robotic Warfare Drawing Nearer », *Washington Times*, 9 février 2005, et repris par PAX, *op. cit.*, p. 1.

23. Comité international de la Croix-Rouge (CICR), *Report of the ICRC Expert Meeting on "Autonomous Weapon Systems"*, 9 mai 2014, p. 12.

24. Voir CICR, *Guide de l'examen de la licéité des nouvelles armes et des nouveaux moyens et méthodes de guerre*, 2006.

Sont illégales les armes qui, par nature, ne peuvent discriminer, c'est-à-dire cibler seulement un objectif militaire, ou causent des maux superflus. C'est le premier risque qui est ici en cause : les SALA peuvent-ils être dirigés contre un objectif militaire seulement, et leurs effets peuvent-ils être limités ? Une réponse négative signifierait qu'ils violent le principe de précaution (PI, art. 57-2).

Les roboticiens exagèrent souvent leur capacité à programmer le DIH, à convertir des règles de droit en algorithmes. Non juristes, ils ont une compréhension simpliste des règles, les réduisant à des commandes univoques : « si c'est un civil, ne tire pas » / « si c'est un combattant, tire ». Outre qu'il est de moins en moins facile, dans les conflits contemporains, de distinguer le civil du combattant, il peut être légal de tirer sur un civil s'il participe directement aux hostilités – une notion complexe qui donne lieu à des interprétations divergentes –, et il peut être illégal de tirer sur un combattant s'il est hors de combat, ce qui n'est pas non plus simple à établir²⁵. L'application du principe de distinction – savoir sur quoi ou qui tirer – dépend du contexte. Celle du principe de proportionnalité est encore plus difficile puisqu'il s'agit de mesurer le caractère éventuellement excessif des dommages collatéraux en fonction de l'avantage militaire attendu, ce qui implique une évaluation au cas par cas, dans un contexte stratégique et politique dont la machine n'a tout simplement pas connaissance.

D'une manière générale, le respect du DIH dépend du contexte et d'un raisonnement de type délibératif, pour lequel les humains sont mieux équipés que les ordinateurs. Il est difficile d'imaginer que ces armes puissent un jour y suffire, à moins qu'un homme, en communication permanente avec la machine, ne les alimente, auquel cas le système d'armes ne serait plus totalement autonome mais sous autonomie supervisée.

Ceci étant dit, les hommes eux-mêmes s'avèrent peu capables de respecter le DIH, souvent violé et suscitant d'interminables controverses doctrinales et judiciaires. Il faut donc rejeter l'argument des opposants aux SALA qui invoquent l'impossibilité de ces derniers non seulement temporaire mais absolue de respecter un jour le DIH : ils exigent l'infailibilité, au lieu d'une faillibilité moindre ou égale à celle des hommes. On peut exiger du système qu'il passe ce que George Lucas appelle le « test d'Arkin »²⁶, adaptation du fameux test de Turing en matière d'intelligence artificielle selon lequel le comportement de la machine doit être indifférenciable du comportement humain dans un contexte donné. Un robot satisfait aux exigences

25. Il peut aussi être pertinent de ne pas tirer sur un combattant pour des raisons tactiques, comme ne pas découvrir sa position par exemple.

26. Du nom du roboticien américain Ronald C. Arkin.

légales et morales – et peut par conséquent être déployé –, lorsque l'on fait la démonstration qu'il peut respecter le droit des conflits armés *aussi bien ou mieux* qu'un humain dans des circonstances similaires²⁷. Ce qu'il faut exiger du système, par exemple, c'est « d'être en mesure de reconnaître les blessés, non pas comme Dieu pourrait le faire, mais comme pourrait le faire un être humain²⁸ ». Rien ne permet de dire que les SALA n'en seront pas capables un jour. On pourrait d'ores et déjà mettre en œuvre un protocole de test dans lequel le système devrait identifier et caractériser des comportements illustrés dans une vidéo, par exemple, et comparer les résultats aux performances humaines.

Le cas échéant, si un système passe ce test, non seulement il peut être déployé mais nous avons même l'obligation morale de le faire²⁹, en vertu de ce que Lucas nomme, après Bradley Jay Strawser³⁰, « le principe du risque non nécessaire » : dans le contexte d'un conflit par ailleurs légalement et moralement justifié, nous avons l'obligation de minimiser les risques encourus par les combattants³¹ – c'est-à-dire de les remplacer ou de les assister avec des machines, à condition qu'elles satisfassent le test d'Arkin.

Aux prévisionnistes – les SALA seront-ils capables de respecter un jour le DIH ? –, se pose la question de savoir si la conformité au DIH exige quelque chose d'inhérent au jugement humain. La particularité des machines par rapport aux humains est qu'elles sont dénuées d'émotions. Les défenseurs des SALA en concluent qu'elles respecteront mieux le DIH

Les émotions : positives ou négatives ?

que les humains puisque, étant dénuées d'instinct de conservation, elles ne seront pas incitées à utiliser une force excessive pour se protéger et, dénuées de stress et de sentiments comme la peur, la vengeance ou la haine, elles commettront moins de crimes. Ne craignant pas de poursuites judiciaires, elles n'auront pas non plus de raison de dissimuler des informations. Le SALA pourrait même, par sa présence dans une équipe humaine, pousser les soldats à mieux respecter le DIH : ses capteurs enregistrant leurs actions, il jouera un rôle de surveillance (*monitoring*) et l'on peut même imaginer qu'il puisse dénoncer directement à la hiérarchie une violation du DIH.

27. Voir par exemple G.R. Lucas, « Automated Warfare », *Stanford Law & Policy Review*, vol. 25, n° 2, 2014, p. 322, 326 et 336.

28. M. Sassòli, réunion de la CCAC citée, Genève, 14 mai 2014.

29. R.C. Arkin, *Governing Lethal Behavior in Autonomous Robots*, Londres, Chapman and Hall / CRC, 2009 et « The Case for Ethical Autonomy in Unmanned Systems », *Journal of Military Ethics*, vol. 9, n° 4, 2010, p. 332.

30. B.J. Strawser, « Moral Predators: The Duty to Employ Uninhabited Vehicles », *Journal of Military Ethics*, vol. 9, n° 4, 2010, p. 342.

31. G.R. Lucas, *op. cit.*, p. 334.

Bien entendu, le commandant humain reste lui soumis à ces émotions, et on ne peut exclure qu'il veuille utiliser un SALA par vengeance, par exemple. Plus généralement, le comportement du système dépendant de sa programmation, et tous les humains n'étant pas bien intentionnés, ses opposants ont envisagé qu'un SALA puisse être programmé pour commettre des crimes de guerre. Contrairement au soldat, qui a en principe une inhibition naturelle à tuer³², qui peut faire preuve de compassion et est doté d'un sens moral qui peut le faire désobéir, la machine les exécuterait sans question. L'argument de l'absence d'émotions est à double tranchant : le SALA est privé des émotions humaines qui peuvent causer des crimes comme de celles qui peuvent les éviter.

Que l'on puisse faire un usage illégal et immoral de cette arme comme de n'importe quelle autre n'est pas un risque suffisant pour l'interdire. On ne cesse pas de produire des avions parce qu'ils peuvent être détournés par des terroristes. « Si les méchants abusent d'une chose – disait Hugo Grotius en 1625 –, le droit ne cesse pas aussitôt pour cela d'exister. Les pirates naviguent aussi, les brigands font aussi usage du fer³³. »

Contre une interdiction préventive

Certains estiment que le droit existant encadre suffisamment les SALA ; d'autres réclament une interdiction préventive – mesure exceptionnelle mais pas unique, celle des armes à laser aveuglantes en 1995 constituant un précédent. Les ONG demandent une telle mesure pour tout système d'arme opérant sans contrôle humain significatif lors des attaques individuelles.

Le caractère préventif de cette interdiction invalide le parallèle que ces organisations développent avec les mines antipersonnel, pour montrer qu'une mobilisation importante de la société civile peut aboutir à une interdiction (convention d'Ottawa, 1997). La différence est pourtant évidente : les mines, qui ont tué et continuent de tuer des millions de civils, avaient fait la démonstration de leur illégalité au regard du DIH (violation du principe de distinction), alors que les SALA n'ont encore rien prouvé, et ne violent *a priori* aucun des principes du DIH. Pour cette raison, estime Marco Sassòli, « une interdiction ne peut pas se fonder sur le DIH³⁴ ».

Ceux qui le reconnaissent invoquent alors le principe de précaution. Mais celui-ci peut tout aussi bien être brandi par les défenseurs des SALA,

32. C'est du moins la thèse du lieutenant-colonel D. Grossman, *On Killing: The Psychological Cost of Learning to Kill in War and Society*, New York, NY, Little, Brown and Co, 2009.

33. H. Grotius, *Le Droit de la guerre et de la paix*, Paris, PUF, 2005, ch. XXV, VIII, 4, p. 568.

34. M. Sassòli, réunion de la CCAC citée.

qui croient que ces machines seront, plus que les humains, capables de respecter le DIH, dont l'un des principes est précisément la précaution de ne pas causer de dommages aux personnes et aux biens. Au devoir de ne pas développer des technologies potentiellement dangereuses s'oppose en effet celui de le faire si celles-ci peuvent réduire l'impact des conflits armés sur ses propres forces et sur les civils. Le cas échéant, il ne serait pas seulement moral de les utiliser, mais immoral de ne pas le faire. Comme à ce stade il est impossible de confirmer ou d'infirmer ces deux hypothèses, et de prédire les développements futurs qui pourraient éclairer la décision, il n'y a pas de raison suffisante de les interdire.

Toute tentative de fonder aujourd'hui une interdiction sur un risque non avéré posé par des armes qui n'existent pas encore est par nature vouée à l'échec. Il n'y a en réalité aucune base sérieuse à une interdiction totale et *a priori*. Quelle que soit sa forme (convention internationale indépendante ou nouveau protocole à la Convention sur les armes classiques de 1980), il faut d'ailleurs souligner les limites de toute voie conventionnelle : d'une part, elle ne s'appliquerait qu'aux États – et seulement aux signataires, qui peuvent émettre des réserves –, et pas aux acteurs non étatiques alors que c'est sans doute d'eux que vient la menace la plus grande, et, d'autre part, il serait difficile de vérifier sur le terrain si la convention est effectivement respectée.

Une interdiction préventive aurait en outre des effets pervers. D'une part, elle ne serait pas respectée – et il serait contreproductif et dangereux pour le droit de construire un régime juridique non respecté. D'autre part, ces technologies étant fortement duales, une interdiction pourrait affecter le marché, pourtant utile et prometteur, des applications civiles.

Mettre en place des garde-fous

Le premier garde-fou est le droit lui-même. Les opposants aux SALA qui présument l'impossibilité pour ces derniers de respecter le DIH mettent en évidence un faux problème. De la même manière qu'il est inutile de craindre des armes pleinement autonomes qui n'existeront pas parce que les humains n'ont pas d'intérêt à les produire, il est inutile de craindre des armes incapables de respecter le DIH car, si c'était le cas, elles n'entreraient simplement pas en service – pour une raison légale (PI, art. 36) et parce qu'il n'est pas dans l'intérêt de la puissance utilisatrice d'avoir une arme incapable de ne viser que des objectifs militaires et qui tue donc trop de civils.

D'autres garde-fous permettent de minimiser les risques liés à l'imprévisibilité. Premièrement, n'utiliser les SALA que contre certains objectifs

militaires. Le DIH distingue entre des objets qui *sont* des objectifs militaires par nature (installations, véhicules, systèmes d'armes, etc.) et d'autres qui le *deviennent* par « leur emplacement, leur destination ou leur utilisation » (PI, art. 52-2). Pour éviter à un SALA d'avoir à faire des choix difficiles comme décider si un bien civil, une ambulance par exemple, a perdu sa protection et est devenu une cible militaire légitime par son emplacement, sa destination ou son utilisation, il suffit de limiter son usage aux objectifs militaires de la première catégorie (par nature). La machine n'a pas besoin de savoir distinguer un civil d'un combattant pour identifier un tank ou une batterie antiaérienne.

Une objection à cette proposition est qu'un objectif militaire par nature peut cesser de l'être par son emplacement : si un tank ou une batterie antiaérienne est dans une cour d'école, par exemple, ou si une colonne de réfugiés passe à proximité. Une fois de plus, la légalité de la cible est contextuelle : le SALA peut être programmé pour détecter l'objet, mais peut-il évaluer son environnement ? La réponse à cette difficulté est simple : il suffit de contrôler le contexte dans lequel l'arme est déployée.

Deuxièmement, donc, n'utiliser les SALA que dans certains contextes. Leurs opposants les imaginent déjà comme des armes terrestres déployées dans des zones habitées, où la distinction entre civils et combattants est délicate. Ils ont alors beau jeu de conclure que les SALA ne seront pas capables de faire cette distinction. Ce faisant, ils ignorent la très forte logique de milieu à laquelle ces systèmes sont soumis : particulièrement adaptés aux univers sous-marin, marin, aérien et spatial, où le risque de cibler accidentellement des civils est faible, ils ont en réalité très peu d'intérêt opérationnel en milieu urbain, précisément parce que le risque de dommages collatéraux compromet l'objectif de gagner les cœurs et les esprits de la population, si important dans les luttes anti-insurrectionnelles.

Pas de SALA en milieu urbain

On renonce déjà à utiliser des armes plus autonomes, donc plus imprévisibles, dans des milieux à forte densité de civils : les *Special Weapons Observation Reconnaissance Detection Systems (SWORDS)*³⁵ déployés en Irak en 2007 ont finalement été retirés en raison d'un comportement dangereux ; les armes amorcées par capteurs sont très peu utilisées ; et le missile *Brimstone* a dû être modifié, à la demande du ministre britannique de la Défense, pour le conformer aux règles d'engagement d'un terrain

35. Robot terrestre mobile pouvant être équipé de diverses armes (mitrailleuse, lance-grenade, arme incendiaire). Il n'est pas autonome : l'ouverture du feu est contrôlée par un opérateur. Son successeur est le *Modular Advanced Armed Robotic System (MAARS)*.

complexe comme l'Afghanistan – il en a été demandé une version guidée par laser dont la cible serait choisie par un opérateur humain³⁶.

Dire que les SALA devraient être interdits parce qu'ils sont ou seront incapables de distinguer un civil d'un combattant, s'ils ne sont pas déployés dans des contextes où ils auront à faire une telle distinction, est un sophisme. Cette incapacité supposée – qui reste encore à prouver – est une raison suffisante de ne pas les déployer dans des zones urbaines, mais pas de les interdire purement et simplement puisque tous les champs de bataille ne comprennent pas des civils ou des objets civils³⁷.

Troisièmement, on peut même prioriser les objectifs militaires. Le DIH exige qu'ils apportent « une contribution effective à l'action militaire » et que leur destruction offre « un avantage militaire précis » (PI, art. 52-2) – deux expressions suscitant des interprétations divergentes : celle, étroite, du CICR contre celle, large, des États-Unis par exemple. Le SALA devra-t-il lui aussi produire sa propre interprétation ? Non, car ces subtilités ne valent que pour les cas limites, les biens à double usage contre lesquels ces systèmes ne devraient de toute façon pas être déployés. Si leur usage est restreint à des objectifs militaires par nature comme une base militaire ou un chasseur bombardier, il est difficile de prétendre que ceux-ci ne contribuent pas effectivement à l'action militaire et que leur destruction n'offre pas d'avantage militaire précis. Il est en outre possible d'introduire une priorisation : le *Low Cost Autonomous Attack System* de Lockheed Martin traite les cibles dans l'ordre de priorité programmé – il peut par exemple identifier les missiles sol-air 9K33 et les tanks T-72 mais ne détruire que les premiers.

Quatrièmement, on peut aussi programmer le bénéfice du doute. Face à un imprévu, le SALA pourrait s'arrêter et consulter sa hiérarchie – application de la règle « en cas de doute, ne tire pas ». Mais l'ennemi pourrait en profiter pour créer des situations imprévues pour paralyser les systèmes – ce qui, dans l'état actuel du DIH, ne constituerait pas nécessairement un acte de perfidie³⁸.

Cinquièmement, l'homme peut aussi conserver la possibilité de désactiver à distance la fonction tir (*veto power*). Cette précaution est conforme à la règle 19 du DIH coutumier – « faire tout ce qui est pratiquement possible

36. N. Marsh, *op. cit.*, p. 3.

37. M. Schmitt, « Autonomous Weapon Systems and International Humanitarian Law: A Reply to Critics », *Harvard National Security Journal*, 2013, p. 11.

38. M. Sassòli, réunion de la CCAC citée.

pour annuler ou suspendre une attaque lorsqu'il apparaît que son objectif n'est pas militaire ou que l'on peut attendre qu'elle cause incidemment des pertes » civiles. Parler d'un mode « sans supervision » reste pertinent s'il signifie l'absence, non de contrôle au sens de capacité de désactiver le mode (*override*), mais de surveillance (*oversight*).

L'idée, derrière ces garde-fous, est la prévisibilité – ce qui est paradoxal puisque les SALA, par définition innovants, pourront avoir des effets non maîtrisables liés à leur autonomie. C'est précisément pour que cette imprévisibilité ne soit pas un problème que cette autonomie ne sera pas totale : ils ne seront déployés que si l'on estime pouvoir prévoir leur comportement.



Mots clés

Systèmes d'armes létaux autonomes
Robots tueurs
Robotisation militaire
Droit international