

L' « autonomie » de la machine dans les systèmes homme-machine : évolution ou révolution du champ de bataille ?

Aspects juridiques et éthiques

Florian Gros⁺ (Onera), Thierry Pichevin^o (CREC Saint-Cyr), Eric Pomès (Cerdes)^{*}, Catherine Tessier^{**} (Onera)

Introduction

Le thème de la robotisation du champ de bataille est d'une grande actualité. Ce thème est par excellence pluridisciplinaire. La présente contribution l'abordera sous un angle limité, en croisant les approches scientifique, éthique et juridique. L'emploi de machines dotées d'une plus ou moins grande autonomie¹ possède en effet d'évidentes implications juridiques et éthiques.

La robotisation du champ de bataille doit être replacée dans le contexte de l'évolution de la conflictualité moderne. Les engagements armés évoluent avec la quasi disparition des conflits interétatiques et la multiplication corrélative des conflits intra ou transétatiques qu'autorise la perte par les Etats du monopole de la violence armée. Le type d'engagement, la diminution des budgets de défense et la volonté d'épargner la vie de soldats expliquent, en partie, le développement des systèmes autonomes. Si cette évolution technologique n'est que le dernier avatar du perfectionnement des armes dans l'histoire militaire, l'utopie technologique souligne cependant la spécificité du phénomène. La croyance que la guerre pourra être exclusivement menée avec des systèmes en se passant d'hommes anime un certain nombre de penseurs. Or la nature de la guerre implique forcément la perte d'êtres humains, ne serait-ce que les cibles de ces systèmes.

+++ Doctorant DGA-Onera sur le sujet *Éthique et partage de l'autorité pour des engins à autonomie variable*.

***° Docteur en science et en éthique, enseignant-chercheur aux Ecoles de Saint-Cyr Coëtquidan.

* Docteur en droit, chercheur associé au CERDES, EA n°3180, Université de Nice Sophia Antipolis et au CREC Saint Cyr, Chargé de cours aux Ecoles de St Cyr Coëtquidan et à l'Université de Nantes.

***** Chercheur sur les questions d'autonomie, de partage de l'autorité, d'éthique liées à la robotique.

¹ On emploiera les termes machine, robot ou système sans équipage (embarqué). Dans la littérature il apparaît que l'utilisation du terme robot devrait être évitée au profit d'autres qualificatifs plus précis comme « système d'arme autonome », « système d'arme autonome humainement supervisé », « système d'arme semi-autonome » (Department of Defense 2012).

Loin de la simple hypothèse utopiste, le déploiement et le développement de ces systèmes est aujourd'hui une réalité confirmée par les conflits irakien et afghan ou la lutte contre le terrorisme. D'innombrables questions se posent face à cette réalité. Le déploiement des systèmes sans équipage embarqué pose la question centrale de l'interaction entre l'homme et la machine. En effet, comment positionner la machine par rapport à l'homme et faire que cette interaction soit claire, compréhensible et prédictible, de façon à ce que son résultat (l'action produite par le système) soit légalement et moralement acceptable ?. La réponse à cette question nécessite d'une part que soit clarifié le concept d'autonomie et d'autre part d'aborder le partage d'autorité et la faisabilité de la programmation des principes du droit international humanitaire. La conception et le déploiement des systèmes sans équipage embarqué montrent que l'autonomie est une contrainte qui rejaillit sur l'attribution des responsabilités en particulier en raison de la complexité d'implémenter les règles du droit international humanitaire.

Les systèmes homme-machine : la contrainte de l'autonomie

Même si la recherche sur l'autonomie robotique existe depuis longtemps (par exemple le « NRL Electric Dog » a été créé en 1923), les avancées technologiques du domaine (Naval Research Laboratory 2011), le déploiement de systèmes robotiques lors des récents conflits (Afghanistan, Israël, frontières de la Demilitarized Zone coréenne (Human Rights Watch 2012)), ainsi que les pressions économiques et militaires pour diminuer le nombre d'opérateurs humains, voire faire superviser plusieurs robots par un seul opérateur (Murphy and Burke 2010), font que les différentes implications de l'autonomie en robotique militaire sont largement étudiées. La question de l'autonomie des robots a donc gagné en importance aussi bien dans l'opinion publique (Heyns 2013) que dans la recherche scientifique, ce qui a entraîné une abondance de définitions techniques, qui viennent s'ajouter aux définitions des domaines philosophiques et juridiques.

Cela étant, même si de nombreuses définitions de l'autonomie robotique sont proposées (certaines seront exposées plus loin), il serait trompeur de penser que dans les prochaines années des robots ayant des capacités décisionnelles et une autonomie similaires à celles des humains puissent être conçus. Les obstacles techniques sont trop importants et, dans le cas des robots armés, des questions de nature éthique et juridique (par exemple, le respect des principes du droit humanitaire, l'attribution de la responsabilité) doivent être prises en compte. Les limitations actuelles rendent donc nécessaire la présence d'opérateurs humains pour superviser les actions du robot lorsqu'il est déployé, et reprendre la main dans des situations particulières. Il semble alors préférable de concevoir l'autonomie non pas pour les robots comme artefacts isolés d'un contexte, mais au sein d'un système constitué du robot et de l'humain.

Dans cette partie nous commencerons par regarder, parmi les conceptions de l'autonomie pour les robots proposées dans la littérature scientifique, quelles sont celles qui peuvent s'appliquer aux systèmes conjoints homme-robot. Nous verrons ensuite l'importance de cette relation dans les aspects de partage du contrôle sur les fonctions du robot. Nous soulignerons enfin les difficultés techniques qui sont rencontrées pour rendre les actions des robots conformes aux principes du droit humanitaire et des lois de la guerre.

L'autonomie, une notion, des définitions

En général, l'autonomie peut être vue comme la capacité de quelque chose à évoluer ou à fonctionner indépendamment de quelque chose d'autre (Castelfranchi and Falcone 2003). Il existe cependant un grand nombre de définitions de l'autonomie dans la littérature ; nous allons voir quelles sont celles qui peuvent s'appliquer aux systèmes homme-robot, en tant qu'elles reflètent mieux la réalité du terrain. En effet, vu que les robots actuellement conçus ne sont pas entièrement autonomes, ils sont nécessairement supervisés par des opérateurs humains et leur « autonomie » ne concerne qu'un certain nombre de fonctions précises. Il peut cependant y avoir une confusion entre ce type d'attribution de l'autonomie et la notion d'automatisme ; il convient donc de distinguer ces deux termes avant de considérer l'autonomie en elle-même.

Automatisme et autonomie

Sans plonger dans les définitions formelles de l'automatisme comme celle qui est proposée par (Castelfranchi and Falcone 2003), il y a plusieurs aspects sur lesquels l'autonomie et l'automatisme diffèrent, à savoir la prévisibilité des actions, la structure de l'environnement et la relation à l'humain. Un automate réalisera pas-à-pas une séquence d'actions déterminée a priori (Truszkowski et al 2009). Dans ce cas, et sauf défaillance, les actions de la machine sont totalement prévisibles et ne peuvent pas être adaptées à un état imprévu de l'environnement : la machine doit donc fonctionner dans un environnement structuré, fixe, pour accomplir avec succès sa séquence d'actions (Human Rights Watch 2012). Cette inflexibilité est due en partie au but des systèmes automatiques, qui est de remplacer les processus manuels de routine par des processus effectués par un dispositif physique et/ou un programme informatique (Truszkowski et al 2009). Même si ces routines peuvent comprendre une participation de l'homme, elles sont conçues pour donner des résultats prévus et désirés par l'homme.

Le robot autonome sera quant à lui capable de fonctionner dans des environnements ouverts et non structurés. C'est la capacité à s'adapter à un environnement variable qui est importante ici. Le robot est toujours contrôlé par un ordinateur mais les programmes sont conçus pour interpréter l'information issue des capteurs, déterminer les actions les plus pertinentes en fonction de cette interprétation et

calculer quand et avec quelles ressources ces actions doivent être effectuées. Si le but est toujours de réaliser des actions sans avoir besoin de l'intervention humaine, l'autonomie cherche à imiter le comportement de l'homme plus qu'à le remplacer (Truskowski et al 2009), à effectuer une action « comme l'homme » plus que « sans l'homme » (Jones 2008). En guise d'illustration, nous pouvons prendre la fameuse distinction entre une machine à laver et un robot en mission de reconnaissance. La machine à laver effectuera toujours les mêmes actions dans le même ordre selon une consigne donnée, de façon à produire un résultat prévisible.

À l'opposé, le robot en reconnaissance devra adapter son comportement à un environnement imprévisible, en réagissant dynamiquement aux événements extérieurs. Cependant l'adaptation restera bornée – les actions sont en effet limitées par l'information dont dispose le robot, le temps disponible pour les calculs et les possibilités des algorithmes (Defense Science Board 2012, par transposition de la rationalité limitée (Simon 1990)).

Autonomie dans le système homme-robot

Pour considérer l'autonomie du robot dans le système homme-robot, nous avons besoin de prendre en compte ce que la machine peut faire, et de considérer ces capacités dans la relation entre elle et l'humain. Les définitions de l'autonomie proposées dans la littérature et qui peuvent s'appliquer sont nombreuses, mais peuvent être regroupées en plusieurs catégories.

La première catégorie est celle de l'autonomie de but et de l'autonomie environnementale. L'autonomie de but est le concept le plus intuitif lorsque l'on aborde l'autonomie. Elle peut se comprendre à la fois comme la capacité à générer ses propres buts et objectifs (Luck et al 2003), mais aussi comme la capacité à agir en accord avec ses propres buts, perceptions, états internes et connaissances sans intervention extérieure (Truskowski et al 2009). Cela étant, outre les problèmes techniques que cet aspect de l'autonomie soulève, ces capacités nécessitent que l'agent comprenne ses propres connaissances et compétences, l'environnement dans lequel il évolue, et surtout comment ses actions peuvent modifier cet environnement pour atteindre le but fixé. Cette compréhension par l'agent de l'interaction entre lui-même et son environnement peut se référer à l'autonomie environnementale, où l'agent peut agir en s'adaptant aux contraintes de son environnement extérieur, ce dernier pouvant être de nature physique ou sociale. L'agent autonome par rapport à son environnement physique sera capable d'anticiper (jusqu'à un certain point) ses actions et leurs conséquences, sur la base des stimuli environnementaux qu'il perçoit ou des forces physiques qu'il subit (Bradshaw et al 2003; Pomès 2012). L'environnement social d'un agent est l'ensemble des agents avec qui ce dernier peut interagir, sachant que les composants de cet environnement peuvent avoir une influence sur l'agent en question.

Dans un environnement social, l'autonomie est la capacité d'agir indépendamment des autres agents et de leur influence ou de leurs préférences (Castelfranchi and Falcone 2003), ainsi que l'absence de supervision dans le cadre d'un système humain-robot (Schreckenghost et al 1998). Cette indépendance est aussi considérée par (Lin, Bekey and Abney 2008) quand la machine opère pour des périodes longues dans le monde réel pour au moins plusieurs domaines d'opération. Ces deux approches de l'autonomie posent une grande quantité de problèmes techniques (que nous aborderons plus loin), mais elles constituent une bonne introduction à des conceptions de l'autonomie dans le cadre des systèmes humain-robot.

La seconde catégorie est celle de l'autonomie juridique. Le point important pour accéder à l'autonomie juridique est la possession d'une compétence, à savoir la capacité physique ou cognitive d'accomplir une tâche donnée. Cette compétence donne aussi à l'individu une autorité sur les autres par rapport à l'objet de la compétence. En effet, si une personne est compétente pour une tâche, alors elle sera à même de juger si ses propres décisions (ou celles des autres) sont appropriées (Daups 2012). De cette façon un individu compétent est capable d'imposer sa décision à d'autres tant que cette compétence est reconnue comme telle. L'individu compétent est alors autonome juridiquement en tant qu'il possède des droits et des devoirs correspondant à la reconnaissance de sa compétence légale. En effet, si une personne a l'autorité sur d'autres individus dans ses décisions, il devient responsable de ses décisions au regard desdits individus. Cette autonomie juridique s'avère pertinente dans le cadre d'un système homme-robot, en tant que les deux agents ont généralement pour but de coopérer en vue d'un but commun. Chaque agent doit alors connaître les compétences de l'autre de façon à se partager les tâches et les accomplir de façon efficace. De la même façon, la compétence peut être un facteur déterminant dans un système où le contrôle est partagé de façon dynamique : tout événement qui change la compétence d'un agent pour une tâche donnée (par exemple une panne, la fatigue) peut déclencher une modification de l'affectation du contrôle de certaines fonctions du système.

L'autonomie exécutive constitue la troisième catégorie. Cette autonomie donne à un agent un niveau de décision et de liberté pour accomplir un objectif qui lui est assigné (Zieba et al 2011). Contrairement à l'autonomie de but, cet agent ne peut pas ou n'est pas autorisé à générer ou à poursuivre ses propres buts. Ceci est particulièrement important dans le contexte militaire, étant donné que la plupart des robots militaires sont conçus comme des outils spécifiques pour accomplir un nombre restreint de tâches et d'objectifs qui leur sont donnés par une source externe, par exemple un opérateur humain ou un acteur dans la chaîne de commandement. On peut aussi considérer l'aspect délégitif de l'autonomie exécutive : étant donné que l'agent robotique est utilisé comme un outil en vue de l'accomplissement d'un objectif assigné par l'homme, on peut considérer l'autonomie comme la capacité d'agir selon les

préférences de l'utilisateur. Comme le proposent (Hexmoor, Castelfranchi and Falcone 2003) : « un appareil est autonome quand il adopte les préférences de l'homme et qu'il agit en accord avec elles »². Un opérateur peut par exemple déléguer à un robot une tâche de détection d'intrus dans un bâtiment, en considérant que les enfants ne sont pas perçus comme des intrus. Si le robot est capable de gérer la tâche en respectant cette préférence, on peut considérer que le robot a réalisé cette tâche de façon autonome. Les préférences de l'utilisateur peuvent alors représenter un cadre qui établit une rationalité limitée pour l'agent (Simon 1990). L'autonomie pourra alors être vue comme la capacité de l'agent à trouver des solutions potentielles, à les évaluer par rapport aux attentes de l'utilisateur, et à choisir la solution la meilleure parmi celles disponibles en fonction de ces attentes (Braynov and Hexmoor 2003 ; Huang et al 2005).

Une autonomie fluctuante

Une dernière approche de l'autonomie est de l'envisager sous l'angle de la collaboration. La collaboration entre un homme et un robot implique une relation entre les deux agents. Cette relation peut soit être vue de façon formelle entre un sujet, un objet et une affectation de l'autonomie (Braynov and Hexmoor 2003), soit comme la position du système conjoint homme-robot sur un continuum de contrôle du système partagé entre les différents agents sur un certain nombre de fonctions (voir plus loin). En conséquence, l'autonomie du robot est un continuum allant de situations où l'homme prend toutes les décisions jusqu'à des situations où un grand nombre de fonctions sont déléguées à l'ordinateur, celui-ci restant supervisé par l'homme.

Pendant longtemps on a trouvé dans la littérature différents concepts de « niveaux d'autonomie », établis selon les rôles respectifs du robot et de l'opérateur ou bien selon les tâches affectées au robot ou à l'opérateur. Il s'agissait de concepts abstraits, uniquement descriptifs : « niveaux » rigides, en nombre limité, de définition parfois douteuse, et sans considération de critères pour passer d'un « niveau » à un autre. Cependant l'autonomie est inscrite dans le cadre d'une relation entre l'agent robotique et l'agent humain (Castelfranchi and Falcone 2003), et cette relation n'est ni uniforme ni figée dans un système donné et pour une mission donnée. Ainsi le Department of Defense américain (Defense Science Board 2012) recommande de considérer l'autonomie comme l'affectation de fonctions cognitives et de responsabilités à l'homme et à la machine pour la mise en œuvre de capacités spécifiques, cette affectation pouvant varier au cours d'une mission.

² Il est à noter qu'adopter des préférences externes n'abolit pas l'autonomie exécutive mais ne fait qu'ajouter certaines contraintes au « comportement » de l'agent, sans définir une séquence fixe d'actions à exécuter. L'agent n'est alors pas considéré comme un automate.

De plus, il existe un lien entre le milieu de déploiement (air, mer, terre) et l'autonomie. En effet, la complexité du milieu, notamment le nombre d'objets, doit être prise en compte. Plus le milieu est simple plus l'autonomie de la machine peut être importante. Au contraire, la complexité du milieu milite davantage pour une réduction de l'autonomie. Ainsi, une autonomie assez large peut être imaginée pour les aéronefs sans équipage embarqué car même s'il faut tenir compte de la circulation aérienne, le milieu aérien apparaît comme relativement simple à gérer. À l'inverse, un robot terrestre déployé en zone urbaine évolue dans un milieu complexe, ce qui peut justifier une part moins grande d'autonomie laissée à la machine. De même, l'autonomie pourrait différer selon le type de conflit. Elle est envisageable dans les conflits de haute intensité dans lesquels l'adversaire est facilement identifiable. En revanche, l'autonomie risque d'être problématique dans les conflits de basse intensité ou de contre insurrection dès lors que les adversaires ne sont pas clairement identifiables.

Enfin, dans le même temps, certaines fonctions peuvent être déléguées (par exemple la navigation d'un drone) tandis que d'autres nécessitent l'action de l'homme (par exemple l'interprétation des images prises par les caméras du drone). L'autonomie n'est donc pas une propriété intrinsèque du robot : la conception et l'opération des engins autonomes doivent être considérées en termes de collaboration homme-machine (Defense Science Board 2012).

L'autonomie, une complexification inévitable de la responsabilité

Le partage des tâches dans un système homme-machine pose un certain nombre de problèmes. Certains relèvent des faiblesses de l'un ou de l'autre. En effet, la machine a des capacités limitées et fait ce pour quoi elle est programmée. Pour sa part, l'opérateur est faillible. Par exemple, il peut se concentrer excessivement sur certains paramètres au détriment des autres – « tunnélisation attentionnelle » (Regis et al. 2012) ; il peut être soumis au phénomène de « Moral Buffer » (Cummins 2006) – distanciation morale par rapport aux actions : ce phénomène peut avoir des conséquences positives (l'opérateur, éloigné du terrain, est moins enclin à agir sous l'effet des émotions) mais aussi négatives (l'opérateur, éloigné du terrain, peut agir sans émotion). D'autres difficultés tiennent aux interactions homme-machine. Deux situations méritent d'être citées : il est possible que se produisent des aléas de communication isolant ainsi le robot de son opérateur ; il peut également se produire le phénomène d'« Automation surprise » (Sarter, Woods and Billings 1997) – rupture dans la conscience de situation : l'homme croit que le robot est dans un certain état alors que la machine est dans un autre état (par exemple parce que des actions ont été effectuées et n'ont pas été notifiées à l'homme, ou bien l'homme n'a pas perçu ces notifications).

Mais plus fondamentalement, la définition de l'autonomie et de son périmètre, comme la place de l'homme dans le processus, nécessitent de s'interroger sur leur impact sur la responsabilité. Nous en évoquerons la nécessité d'un point de vue éthique, avant d'en aborder en détail la faisabilité d'un point de vue juridique.

Quelques considérations éthiques

La nécessité de cerner la responsabilité morale

La question de la responsabilité dans l'emploi de systèmes homme-machine peut être envisagée par rapport à d'éventuels dommages que causerait le système, et aux poursuites qui pourraient en résulter – nous l'aborderons un peu plus loin. Mais, même en l'absence de dommages, elle nous semble revêtir une dimension morale fondamentale, qui relève d'un constat finalement assez simple : lorsqu'on ne porte pas le poids de la responsabilité de ses actes, on est moins enclin à agir de façon réfléchie et mesurée. Dans le contexte militaire, Robert Sparrow (Sparrow, 2007) développe plusieurs arguments qui vont dans ce sens. Le premier est kantien : une expression minimale de respect envers l'ennemi est que quelqu'un puisse être tenu responsable de la décision de lui prendre la vie. Ne pas pouvoir le faire reviendrait à traiter l'ennemi « comme de la vermine », sans aucune considération morale. Cet argument concerne d'ailleurs également les proches des ennemis tués, qui méritent tout autant considération. Le second argument est conséquentialiste : si les responsables de crimes de guerre ne sont pas identifiés, ils ne pourront être condamnés, et cela aura des conséquences désastreuses sur la façon dont les guerres seront menées par la suite.

Les belligérants pourront se croire tout permis, et commettre les actes les plus atroces en toute impunité. Enfin, il s'agit aussi d'une nécessité pour que les principes du Jus in bello (principe de discrimination, principe de proportionnalité) puissent être respectés. En ce qui concerne le cas plus spécifique des drones autonomes, Dominique Lambert (Lambert 2011) montre toute la difficulté à résoudre la question de la responsabilité, et en fait un argument important contre leur usage. Si donc, d'un point de vue éthique, il est indispensable de cerner la question de la responsabilité, comment cette question peut-elle être abordée ? Nous allons voir qu'elle est au carrefour d'un certain nombre de difficultés.

Autonomie et responsabilité morale : une approche délicate

De façon très générale, être responsable signifie, étymologiquement, « répondre de ses actes », c'est-à-dire les assumer, s'en reconnaître l'auteur (Clément et al., 2000). Ce concept s'articule directement à celui de la liberté, car être responsable d'un acte suppose l'avoir effectué librement. Si on m'a contraint à faire quelque chose que je ne voulais pas faire, ou si j'étais sous l'influence d'un désordre mental qui m'enlevait toute capacité de choix, je ne puis être moralement responsable. Inversement, être libre entraîne la responsabilité des actes effectués. Aborder la question de la responsabilité entraîne donc directement à se poser la question de la liberté.

La liberté de l'homme peut bien sûr être questionnée. Elle fait l'objet de nombreuses réflexions philosophiques qui dépassent le cadre de cet article, mais il est intéressant de noter que l'irruption des « robots » conduit à se poser des questions sur ce qu'est l'être humain, quelles sont ses spécificités.

Nous nous contenterons de nous intéresser à la machine, et nous sommes ainsi conduits à nous demander dans quelle mesure la machine est « libre » de faire ce qu'elle fait. Notons tout d'abord que, dans les nombreux travaux portant sur ce thème, ce n'est pas le concept de liberté qui est utilisé, mais celui d'autonomie. D'un point de vue philosophique, ces deux notions ont-elles un lien ? Etymologiquement, autonomie signifie « qui se gouverne par ses propres lois ». La notion de liberté est donc bien présente dans ce concept, puisque « se gouverner par ses propres lois » suppose une indépendance par rapport aux contraintes extérieures. Mais on trouve dans le concept d'autonomie un peu plus que la seule liberté : il y a aussi la faculté de se donner ses propres lois. C'est probablement Kant (Kant 1785) qui a exploré le plus en détail ce concept : l'autonomie kantienne est la capacité d'un individu à créer des lois et des règles auxquelles il soumet volontairement ses propres actions. Cette définition philosophique de l'autonomie peut-elle s'appliquer à une machine ? Une machine est-elle indépendante des contraintes extérieures et peut-elle se donner ses propres lois ? Ces questions nécessitent de s'interroger sur la manière dont la machine est programmée. N'est-elle pas régie par un logiciel, qui lui dit que faire ? Dans quelle mesure les techniques de programmation influent-elles ? Par exemple, une programmation de type « top-down », qui définit à l'avance les actions à mener et les règles à suivre, semble très différente d'une programmation de type « bottom-up », dans laquelle les règles de comportement ne sont pas implémentées à l'avance, mais « apprises » par la machine. Y a-t-il une différence d'autonomie dans ces deux cas ? Ces questions font l'objet de nombreuses recherches, et on pourra consulter la synthèse de Wendell Wallach (Wallach, 2010) pour plus de détails. Elles n'ont pour l'heure pas de réponses précises.

Par ailleurs, nous l'avons vu plus haut, dans le domaine de la robotique, de nombreuses études apportent des définitions de l'autonomie qui diffèrent de celle assez classique rapportée ci-dessus, et qui diffèrent entre elles : autonomie de but, autonomie environnementale, autonomie exécutive, etc. Toutes ces définitions montrent combien la notion d'autonomie de la machine est complexe, et rendent l'abord du concept de responsabilité bien délicat. Il nous semble donc qu'appréhender la responsabilité dans le cadre d'une interaction homme-machine en utilisant la notion d'autonomie est la source de nombreuses complications.

Partage d'autorité et responsabilité morale

Pour sortir de cet imbroglio, il nous semble que recourir à la notion d'autorité (Mercier, Tessier, Dehais 2010) apporte un éclairage intéressant. Un agent (l'homme ou la machine) a l'autorité sur une ressource (une ressource physique, une fonction, une tâche, un but) par rapport à l'autre agent s'il peut contrôler cette ressource, au détriment de l'autre agent. Le partage d'autorité est en général défini lors de la préparation de la mission, mais il peut évoluer au cours de la mission. Cela peut d'ailleurs conduire à des conflits d'autorité, situations dans lesquelles à la fois l'homme et le robot « veulent » contrôler la même ressource (par exemple, l'homme veut examiner précisément une cible en pilotant finement le drone tandis que le drone, qui a détecté une insuffisance de batterie, est programmé pour rentrer à la base) (Tessier and Dehais 2012 ; Pizziol, Tessier and Dehais 2012).

Que pouvons-nous tirer de ce recours au concept d'autorité en ce qui concerne la responsabilité ? Pour répondre à cette question, il convient de préciser la notion de responsabilité (Pichevin, 2012). Si, comme nous l'avons vu, être responsable signifie de façon très générale « répondre de ses actes », il semble utile, pour être plus précis, de distinguer deux concepts de responsabilité (Neuberg 2004). Le premier concept, que nous appellerons la responsabilité-imputabilité, est l'imputabilité des actes effectués : l'agent est responsable des actes dont il est l'auteur. C'est un concept qui nous est relativement familier : l'agent qui casse un vase est responsable de cette détérioration, et donc, par exemple, des réparations à effectuer. Le deuxième concept de responsabilité, que nous appellerons la responsabilité-devoir, provient du devoir acquis du fait de la position occupée : l'agent disposant d'un statut, occupant une fonction, ayant un savoir, peut anticiper les conséquences afférentes. Il a alors le devoir de les prévenir. Ainsi, si ce deuxième agent savait que le vase avait un pied fragile, il porte aussi une part de responsabilité de sa détérioration, car il aurait pu prévenir le premier agent, mettre le vase de côté, etc.

Dans le contexte de notre système homme-machine, la notion d'autorité vient immédiatement faire appel à la responsabilité-imputabilité : est responsable l'agent qui a commis l'acte. Cette définition semble assez simple a priori en ce qui concerne l'homme, mais elle pose question en ce qui concerne la machine : peut-on attribuer une responsabilité à cette dernière ? Cette question, cruciale pour penser la responsabilité de la machine dont nous avons montré la nécessité éthique, sera discutée en détails d'un point de vue juridique plus loin. Au préalable, remarquons à nouveau que la notion de responsabilité-imputabilité est en lien étroit avec la façon dont la machine est programmée. Si c'est avec un algorithme séquentiel classique (« si... alors... », programmation « top-down »), la machine ne fait que suivre la séquence prévue par le programmeur, et la responsabilité-imputabilité de ce dernier est donc directement impliquée. Si la machine est programmée au travers d'un système d'apprentissage (programmation « bottom-up » type réseau de neurones), on pourrait considérer à première vue que la machine « échappe » à son programmeur. Pourtant, la question paraît plus complexe : l'humain qui a procédé à l'apprentissage du robot n'a-t-il pas fait en sorte que l'apprentissage soit orienté dans telle ou telle direction ? Ultiment, les algorithmes d'apprentissage ne sont-ils pas des systèmes écrits par des humains, fonctionnant à base de pondérations qu'ils ont choisies ? Discuter de toutes ces questions dépasserait le cadre de cet article, mais on voit qu'il y a là matière à réflexion.

La notion de responsabilité-devoir introduit une autre perspective. Au-delà de l'acte commis, possède une responsabilité morale celui qui, par sa position, peut l'anticiper, au moins partiellement. Ainsi, toute personne qui, par sa position, peut contribuer à anticiper une action possède la responsabilité de le faire – même s'il est évident qu'on ne peut pas tout prévoir. On voit ici apparaître la responsabilité possible de l'homme « co-pilote » de la machine, même au cas où c'est la machine qui avait l'autorité. On voit aussi apparaître toute une chaîne de responsables, allant du supérieur hiérarchique, en passant par l'acheteur jusqu'au programmeur. En ce qui concerne ce dernier, la responsabilité-devoir peut probablement clarifier la réflexion, notamment lorsque des algorithmes de type apprentissage sont concernés : le programmeur possède au minimum une telle responsabilité à leur égard.

Aspects juridiques

L'analyse qui précède a montré la nécessité morale de cerner les responsabilités lors de la mise en œuvre d'un équipage homme-machine. Elle en a esquissé les contours, mais sa mise en œuvre concrète doit être discutée.

La mise en œuvre des systèmes sans équipage embarqué peut conduire à rechercher la responsabilité civile ou pénale lorsque des dommages ou des violations du droit international humanitaire découlent de leur utilisation. La responsabilité civile vise à la réparation des dommages provoqués, par un fait

quelconque de l'homme source de dommage. Cette forme de responsabilité ne sera pas développée dans cette contribution car elle diffère selon les systèmes nationaux. La responsabilité internationale de l'Etat ne sera pas non plus développée ; elle est cependant envisageable si l'utilisation des systèmes sans équipage embarqué entraîne une violation d'une obligation internationale de l'Etat constitutif d'un fait illicite et imputable à cet Etat³. L'étude abordera davantage la dimension pénale de la responsabilité pour des violations du droit international humanitaire.

L'improbable responsabilité de la machine

L'engagement de la responsabilité pénale d'une personne nécessite la réunion de deux éléments : l'élément matériel, à savoir, le comportement criminel (*actus reus*) et l'élément interne ou mental, à savoir, la connaissance ou l'intention générale vis-à-vis de la conduite (*mens rea*). L'élément psychologique connaît plusieurs niveaux qui vont de la connaissance à la négligence en passant dans certains cas par l'intention spécifique. L'article 30 (1) du Statut de Rome énonce ainsi que « [s]auf disposition contraire, nul n'est pénalement responsable et ne peut être puni à raison d'un crime relevant de la compétence de la Cour [Pénale Internationale] que si l'élément matériel du crime est commis avec intention et connaissance », à savoir lorsque la personne est « consciente qu'une circonstance existe ou qu'une conséquence adviendra dans le cours normal des événements ou lorsque la personne entend adopter le comportement en question et entend causer la conséquence en question ou est consciente que celle-ci adviendra dans le cours normal des événements ». L'article 30 exige donc un élément intentionnel de la part du suspect. Cet élément intentionnel couvre trois hypothèses. Le *dol* direct de premier degré, c'est-à-dire les situations dans lesquelles le suspect sait que ses actions ou omissions seront à l'origine des éléments objectifs du crime et commet ces actions ou omissions dans l'intention concrète de provoquer les éléments objectifs du crime. Le *dol* direct de deuxième degré qui recouvre les situations dans lesquelles le suspect, sans avoir l'intention concrète de provoquer les éléments objectifs du crime, a conscience que de tels éléments résulteront nécessairement de ses actions ou omissions. Le *dol* éventuel renvoie aux situations dans lesquelles le suspect est conscient du risque que les éléments objectifs du crime résultent de ses actions ou omissions et accepte ce résultat en s'y résignant ou en l'admettant.

³ Par exemple, une question concerne la responsabilité en cas de violations de règles du droit international et sur sa conformité au regard de l'article 36 du Premier Protocole additionnel aux Conventions de Genève de 1977. L'article 36 du PA I énonce que « dans l'étude, la mise au point, l'acquisition ou l'adoption d'une nouvelle arme, de nouveaux moyens ou d'une nouvelle méthode de guerre, une Haute Partie contractante a l'obligation de déterminer si l'emploi en serait interdit, dans certaines circonstances ou en toutes circonstances, par les dispositions du présent Protocole ou par toute autre règle du droit international applicable à cette Haute Partie contractante ». Cet article a pour objectif d'obliger les Etats à déterminer la licéité des armes avant qu'elles ne soient mises au point ou acquises. Le développement des systèmes sans équipage embarqué nécessitera donc d'établir si ces systèmes, lorsqu'ils pourront être qualifiés d'armes, ne contreviennent pas à une obligation internationale liant l'Etat.

Pour cette hypothèse la Chambre Préliminaire I de la Cour Pénale internationale a distingué, dans l'affaire Le Procureur C. Thomas Lubanga Dyilo, deux cas de figure. Premièrement, « si le risque de causer les éléments objectifs du crime est élevé (c'est-à-dire, lorsqu'il est probable que cette conséquence « adviendra dans le cours normal des événements »), le fait que le suspect admette l'idée de provoquer les éléments objectifs du crime peut se déduire de ce que :

i. le suspect a conscience de la probabilité importante que ses actions ou omissions se traduiraient par la réalisation des éléments objectifs du crime ; et

ii. malgré cette conscience, le suspect décide de commettre ses actions ou omissions ».

Deuxièmement, « si le risque de causer les éléments objectifs du crime est faible, le suspect doit avoir manifestement ou expressément accepté l'idée que ces éléments objectifs puissent résulter de ses actes ou omissions ». Par conséquent, comme le reconnaît la Chambre Préliminaire dans cette affaire « lorsque l'état d'esprit du suspect ne va pas jusqu'à admettre que les éléments objectifs du crime puissent résulter de ses actes ou omissions, un tel état d'esprit ne saurait être considéré comme une commission véritablement intentionnelle de ces éléments objectifs, et ne remplirait donc pas la condition d' « intention et connaissance » énoncée à l'article 30 du Statut (Le Procureur C. Thomas Lubanga Dyilo, 2007, par. 353-355).

L'élément matériel consiste quant à lui en des actes ou des omissions érigés en infractions pénales. La responsabilité pénale est ainsi engagée soit pour la commission matérielle de l'acte c'est-à-dire par action (acte positif ou omission) soit pour une participation accessoire (complicité). Par exemple, le supérieur hiérarchique peut être tenu pénalement responsable des crimes commis par ses subordonnés (article 28 du Statut de Rome). Il est responsable pour avoir omis de prévenir ou de punir les crimes commis par ses subordonnés. Le supérieur hiérarchique peut ainsi être tenu responsable du comportement prohibé de ses subordonnés pour avoir manqué à son devoir de prévenir ou réprimer leur comportement illégal ou d'en référer aux autorités compétentes. Pour cela, il devait exercer un commandement et un contrôle effectifs, ou une autorité et un contrôle effectifs, sur ses subordonnés qui ont commis le ou les crimes. Ces crimes doivent résulter du fait que le suspect n'ait pas exercé sur ses subordonnés le contrôle qui convenait. En outre, la responsabilité du chef militaire ne sera possible que s'il savait ou, en raison des circonstances, aurait dû savoir que ses subordonnés commettaient ou allaient commettre une ou plusieurs violations du droit international humanitaire. Cela signifie qu'il devait savoir ou aurait dû savoir que ses forces allaient adopter, adoptaient ou avaient adopté un comportement constituant les crimes. Deux situations peuvent être distinguées. La première, recouvrant le terme « savait », impose l'existence d'une connaissance effective. La seconde, renvoie à l'expression « aurait dû savoir », constitue une forme de négligence. Ce qui engage la responsabilité du supérieur est d'avoir négligé de se

renseigner sur le comportement illégal de ses subordonnés. Enfin, il ne doit pas avoir pris les mesures nécessaires et raisonnables qui étaient en son pouvoir pour empêcher ou réprimer l'exécution de ces crimes ou pour en référer aux autorités compétentes aux fins d'enquête et de poursuites (Le Procureur c. Jean-Pierre Bemba Gombo, 2009, par. 402 sq.).

Par exemple, un enfant s'étant approché, lors d'un jeu, de la caserne avec une arme factice a été pris pour cible par la machine chargée de la surveillance qui a reconnue l'arme comme étant réelle et a par conséquent identifié l'enfant comme une menace. Plusieurs solutions sont envisageables. La définition maximaliste de l'autonomie dans laquelle la machine prendrait des décisions sans intervention de l'opérateur nécessiterait de repenser le régime de responsabilité puisque la machine, c'est-à-dire l'objet, serait à l'origine du dommage et de la violation des normes. En effet, ni l'opérateur ni le programmeur ne commettent l'actus reus puisque c'est la machine qui commet l'acte répréhensible, ce qui impose de se demander si la responsabilité de l'un ou de l'autre peut être engagée. Deux hypothèses peuvent être distinguées. Dans la première, la machine n'est que l'instrument permettant la violation ; la décision de la mission et donc des actions sont prises ou initiées par des humains. Il s'agit d'une hypothèse dans laquelle les capacités de la machine ne sont pas véritablement utilisées ; elle n'est qu'un prolongement de l'opérateur. L'acte est donc attribuable à l'opérateur. Une telle hypothèse de commission indirecte se trouve à l'article 25 (3) a) « [a] ux termes du présent Statut, une personne est pénalement responsable et peut être punie pour un crime relevant de la compétence de la Cour si elle commet un tel crime, [...] par l'intermédiaire d'une autre ». La deuxième hypothèse est celle de la machine autonome qui « décide » seule des actions à mener en fonction de son environnement. L'opérateur et le programmeur n'ont ni participé à la réalisation de l'acte ni eu l'intention de commettre l'infraction. La responsabilité de la machine est-elle envisageable ? Cela impliquerait de la reconnaître responsable civilement ou pénalement. Une telle reconnaissance nécessiterait au préalable de repenser la nature même de la machine pour s'interroger sur la pertinence de lui octroyer une personnalité juridique même amoindrie.

Deux arguments militent en faveur du rejet d'une telle responsabilité. D'abord, la machine n'est pas une personne mais une chose⁴, elle ne peut donc être responsable. Une telle reconnaissance n'aurait d'ailleurs que peu de sens puisque les conséquences de la reconnaissance d'une telle responsabilité resteraient plutôt théoriques. La machine serait débranchée ou détruite ce qui équivaldrait à une peine de mort. N'ayant pas de patrimoine, elle ne pourrait pas réparer les dommages. Finalement la conséquence la plus probable serait une reprogrammation. Cela démontre que, même dans le cas d'une autonomie totale (de déplacement et de décision), l'humain reste présent. En outre, l'article 25 (1) du Statut de Rome relatif à la responsabilité pénale individuelle précise que la Cour n'est compétente qu'à

⁴ Le système robotisé ayant une réalité matérielle et étant susceptible d'appropriation est une chose qui selon les situations sera un bien meuble ou immeuble.

l'égard des personnes physiques. Deuxième argument, la machine ne dispose pas de libre arbitre, elle ne fait qu'exécuter les actions selon son programme, par conséquent l'élément intentionnel fait défaut. La pluralité d'acteurs dans la commission de l'infraction commande toutefois que soit envisagée l'hypothèse de la coaction, c'est-à-dire l'accomplissement d'un crime par plusieurs personnes dont chacune est considérée comme l'auteur principal car le comportement de chacun constitue une partie du crime.

Cette hypothèse doit cependant être repoussée car il est impossible de démontrer l'existence d'un accord implicite et a fortiori explicite entre l'opérateur et la machine dans la commission du crime (Le Procureur c. Germain Katanga et Mathieu Ngudjolo Chui, 2008, par. 522-523). Aussi, quelle que soit l'autonomie de la machine, la responsabilité d'un humain devra être recherchée.

Une détermination complexe des responsabilités humaines

La réponse à cette question dépendra là encore de la définition de l'autonomie qui sera retenue. En effet, selon son étendue, le responsable ne sera pas forcément identique. Le responsable pourra en effet selon les cas être l'opérateur (militaire ou civil) lié à la machine, le chef militaire ayant donné l'ordre de déploiement de la machine ou encore le concepteur du logiciel. Toutefois, dans toutes les hypothèses où la machine décide seule, retenir la responsabilité pour crime de guerre de l'opérateur, de son supérieur ou du programmeur semble difficilement envisageable dès lors que l'élément intentionnel fera défaut. Ainsi, par exemple, en cas d'attaque par la machine de personnes civiles ne participant pas directement aux hostilités, l'article 8 (2) b) i) du Statut de Rome exige que l'auteur ait entendu prendre pour cible de son attaque ces personnes civiles ne participant pas directement aux hostilités. Aucun des êtres humains n'ont voulu, ni eu un comportement prouvant l'intention de commettre ce crime. Deux voies s'ouvrent alors. La première est celle de la négligence. L'un et/ou l'autre peuvent voir leur responsabilité engagée car ils auraient dû prévoir ou connaître la commission de l'acte ; l'acte était prévisible. En d'autres termes, ils auraient dû surveiller pour l'opérateur, mieux programmer pour le programmeur. La deuxième voie est celle du crime d'omission. Celui-ci s'entend soit comme la violation d'un devoir d'agir, soit comme une commission par omission, c'est-à-dire une infraction dans laquelle les dommages se réalisent en raison d'une absence d'action. Prenons l'exemple d'un robot terrestre dont la mission serait la surveillance d'une caserne de casques bleus dans une opération de maintien de la paix. Ce système peut fonctionner selon deux modes : la machine prend toutes les décisions même celle de l'application de la force létale, ou la machine fonctionne seule, hormis la décision de recourir à la force armée qui relève de l'opérateur humain. Qui sera responsable si la machine en mode fonctionnement autonome exécute une action qualifiable de crime de guerre ?

Plus sûrement, ce sont des individus qui devront répondre de cet acte. Mais là encore plusieurs possibilités s'offrent à la réflexion. Les responsables peuvent être imaginés en partant du plus près de la machine au plus loin. En premier lieu, l'opérateur humain en charge de la surveillance de la machine pourrait voir sa responsabilité engagée pour ne pas avoir empêché l'acte. La responsabilité de l'individu, militaire ou civil, ayant pris la décision de choisir le mode autonome plutôt que l'autre mode, pourrait être ensuite envisagée. Mais c'est aussi la responsabilité du supérieur hiérarchique, du concepteur du logiciel etc. qui pourrait être mise en cause. La chaîne de responsabilité pourrait ainsi être remontée jusqu'à celui qui a conçu la machine en passant par celui qui l'a testée et/ou certifiée.

Cela démontre que la détermination des responsabilités se fera *in concreto* et non *in abstracto* une fois pour toute. En outre, en cas de pluralité de responsables, les responsabilités ne seront pas alternatives mais cumulatives. Enfin, il est également possible d'imaginer l'intrusion dans la machine de « pirates » pour en altérer sa programmation, ce qui devra avoir des conséquences sur la détermination des responsables. Or, la difficulté d'établir des preuves de tels piratages dans le cyberspace laisse imaginer les difficultés qui se poseront aux juges au moment de déterminer le responsable d'un dommage ou de condamner un individu pour les « faits » d'une machine.

Le partage de l'autorité, élément de complexification de la responsabilité

Plusieurs hypothèses sont alors envisageables, chacune ayant des répercussions sur la responsabilité. L'hypothèse la plus simple est celle de la machine soumise à un contrôle strict de l'opérateur qui prend l'ensemble des décisions. Cette première hypothèse ne pose a priori pas de difficulté : ce sont l'opérateur et le donneur d'ordre qui seront responsables en cas de violation d'une norme juridique. À l'autre extrémité du spectre se trouve l'hypothèse, pour l'instant peu probable, d'une machine qui prend l'ensemble des décisions y compris le recours à la force armée. Entre ces deux hypothèses se trouve l'hypothèse la plus complexe et la plus vraisemblable d'un partage d'autorité selon les tâches ou les missions. Par exemple le Talon SWORDS se déplace de manière autonome mais l'utilisation des armes dépend de l'opérateur. L'établissement de la responsabilité demandera alors une analyse fine de qui avait l'autorité pour décider. Cette hypothèse se complexifie encore avec la possible « reprise en main », c'est-à-dire un changement total ou partiel forcé de l'autorité. Cette « reprise en main » peut être aussi bien le fait de l'opérateur que de la machine. Les cas de responsabilité se multiplient alors. En effet, alors que, selon les spécifications, la machine possède l'autorité, la programmation permet à l'opérateur de reprendre la main. Deux cas sont alors à distinguer. Dans le premier, l'opérateur reprend la main dans des conditions imprévues ou en ne respectant les procédures, pour « forcer » la machine à agir à l'encontre de sa programmation. Clairement, en cas de violation du droit international humanitaire, la responsabilité de l'opérateur et ou du donneur d'ordre sera alors engagée.

Deuxième cas, l'opérateur ne reprend pas la main soit dans des circonstances où il est prévu qu'il le fasse soit alors que les capteurs de la machine l'informe d'une violation imminente du droit. Il s'agit du phénomène d' « Automation Bias » (Cummings 2006) – sur-confiance dans les automatismes : l'opérateur a tendance à s'en remettre aux « décisions » de la machine. Sa responsabilité semble également être engagée pour ne pas avoir agi.

La reprise en main peut être également le fait de la machine. L'hypothèse est celle dans laquelle la machine reprend la main à l'homme et cette reprise en main conduit à de mauvaises interprétations qui donnent lieu à des dommages ou à des crimes. Toute la difficulté réside dans le fait que la machine s'est imposée à l'opérateur. Cette hypothèse peut-elle constituer une cause exonératoire de la responsabilité de l'opérateur et du donneur d'ordre ? En cas de réponse affirmative est-ce que seule la responsabilité de ceux qui l'ont conçue pourra être recherchée ?

L'établissement de la responsabilité nécessite au préalable que soient déterminées les raisons pour lesquelles des erreurs se sont produites. Première hypothèse, la reprise en main est le résultat d'une malfaçon ou des limites du logiciel. La responsabilité sera donc à rechercher du côté des concepteurs du logiciel. La nature de la responsabilité supposera toutefois que soit établi avec certitude le fait que les concepteurs avaient connaissance de la défaillance. En cas d'ignorance, la voie la plus probable est l'engagement de leur responsabilité civile. À l'inverse, s'ils ont commercialisé le logiciel en connaissant la faille, leur responsabilité pénale pourra être recherchée. La deuxième hypothèse est celle d'une perception, d'une interprétation défaillantes des scènes perçues par la machine en raison de la complexité de l'environnement. Toute la difficulté pour déterminer les responsabilités repose alors sur le fait que les résultats dommageables des actions sont dus à une panne ou aux aléas de l'environnement. Faute d'intention, la responsabilité pénale internationale de l'opérateur semble dans ce cas difficile à établir. En droit français, si la victime souhaite engager la responsabilité pénale d'un individu, la voie des délits non intentionnels pourrait être explorée. L'article 121-3 al 4 du code pénal pose en effet que « Dans le cas prévu par l'alinéa qui précède les personnes physiques qui n'ont pas causé directement le dommage mais ont contribué à créer la situation qui a permis la réalisation du dommage ou qui n'ont pas pris les mesures permettant de l'éviter sont responsables pénalement s'il est établi qu'elles ont soit violé de façon manifestement délibérée une obligation particulière de prudence ou de sécurité prévue par la loi ou le règlement, soit commis une faute caractérisée et qui exposait autrui à un risque d'une particulière gravité qu'elles ne pouvait ignorer ». Il s'agit des cas dans lesquels il existe une causalité indirecte, c'est-à-dire une situation où une personne n'a pas créé directement le dommage mais qui a soit créé ou contribué à créer la situation qui a permis la réalisation du dommage, soit n'a pas pris les mesures nécessaires permettant de l'éviter. Pour ce qui est des militaires, l'article L. 4123-11 Code de la défense nationale exige, pour engager leur responsabilité, que soit démontré qu'ils n'ont pas accompli

les diligences normales compte tenu de leur compétence, de leurs pouvoirs et des moyens dont ils disposaient et des difficultés de la mission. Ces critères objectifs invitent le juge à procéder à une appréciation in concreto du comportement du militaire au regard de la spécificité du métier des armes et de la situation particulière et non en l'identifiant à l'homme ordinaire. L'article L. 4123-11 du Code de la défense invite ainsi le juge à apprécier le comportement du militaire en décelant dans la conduite des missions ou des fonctions exercées, dans l'exercice des compétences attribuées, ainsi que dans l'usage des pouvoirs et des moyens dévolus, tous les signes de diligences normales interdisant de considérer la faute comme pénalement constituée, ou au contraire, ceux d'une insuffisance. Le tribunal ne devra pas, cependant, définir ce qu'auraient dû être les fonctions et la mission du militaire mais vérifier s'il a accompli les diligences normales au regard des circonstances de l'espèce. C'est dire que le juge devra rechercher si le comportement reproché au militaire poursuivi est rattachable au dommage en raison d'une insuffisance dans la conduite de ses missions ou dans l'exercice de ses pouvoirs et compétences.

La normalité des diligences pourra, dès lors, être déduite en comparant ces facultés aux initiatives effectivement prises. Il faudra déterminer ce qu'un officier moyen aurait fait dans une telle situation et comparer le résultat de ces investigations avec les comportements du ou des prévenus. Dans le cas d'un dommage causé par la machine il pourrait, par exemple, être reproché au militaire de ne pas avoir tenu suffisamment compte de la météo comme facteur d'erreur pour la machine. L'opérateur aurait ainsi dû savoir qu'au regard du temps le jour des dommages, l'erreur était prévisible. Il est également possible d'envisager la responsabilité de l'Etat devant les tribunaux administratifs.

La délicate régulation du comportement d'un système autonome

L'autonomie de la machine, quelle que soit son étendue, pose la question de la licéité des actes qu'elle va accomplir : comment faire en sorte que les actions résultantes soient acceptables moralement ? Précisons tout d'abord ce que nous entendons par là (Pichevin(2), 2012).

Lorsqu'il faut prendre une décision morale – de l'ordre du bien et du mal –, il est nécessaire de tenir compte de différents types de normes : lois, règles morales. Si la situation est complexe, il faut arbitrer entre ces différentes normes, voire inventer en cas d'absence de normes : on recourt alors ce que l'on peut appeler l'éthique.

Si le terme de loi est courant, il semble important de bien distinguer la morale de l'éthique. La morale est un ensemble de règles de conduite, acceptées plus ou moins implicitement dans une société donnée. Une règle morale typique est « ne pas mentir ». L'éthique désignera la « science » qui consiste à réfléchir et décider, au cas par cas, ce qu'il convient de faire, c'est l'art de résoudre les dilemmes compliqués.

Ces dilemmes se présentent quand des lois, des règles morales, entrent en conflit, ou encore lorsqu'il n'y a aucune règle pour nous guider. Par exemple, une règle morale enjoint de ne pas mentir, mais y a-t-il des situations où le mensonge est néanmoins moralement justifié ? Peut-on mentir pour sauver une personne poursuivie par des assassins ?

Tenter de faire en sorte qu'une machine plus ou moins autonome ait un comportement moralement acceptable nécessite alors d'examiner deux aspects différents : peut-on lui implémenter des règles, et en particulier juridiques ? Et peut-on attendre d'elle qu'elle ait une capacité de réflexion éthique ?

Une implémentation complexe des principes de droit international humanitaire

L'autonomie de la machine peut donc conduire à tenter de programmer les règles juridiques pertinentes (droit international humanitaire, droit de l'espace de déploiement (droit de la mer – par exemple Convention on the International Regulations for Preventing Collisions at Sea, 1972...)) Or, la nature de la plupart de ces règles, en particulier de droit international humanitaire, rend complexe la programmation et constitue peut-être une limite aux missions qui pourraient être confiées aux systèmes homme-robot.

Certes, l'homme ne possède pas forcément les capacités pour appliquer avec cohérence et régularité les principes du droit international humanitaire. Il ne possède pas non plus toutes les connaissances nécessaires pour comprendre intégralement la situation dans laquelle il se trouve et ainsi prendre une décision absolument parfaite. En un sens, l'humain possède aussi une rationalité limitée, qui s'avère juste être moins limitée que celle d'un robot. Certaines émotions comme la peur ou la colère peuvent faire que l'homme ne prenne pas en compte les principes du droit international humanitaire et agisse de manière inacceptable (utilisation disproportionnée de la force, viol, torture...) Un opérateur humain peut subir le phénomène de « Moral Buffer » exposé précédemment, et agir ainsi sans tenir compte des conséquences de ses actions ou des contraintes morales qui lui sont imposées.

Les différents principes qui composent le droit international humanitaire nécessitent néanmoins des prérequis actuellement très difficiles à atteindre, que ce soit pour leur application ou leur implémentation. En outre, comme tout principe général de conduite, ils possèdent deux difficultés inhérentes qui concernent à la fois les hommes et les systèmes conjoints homme-machine. D'abord, ces principes étant fixes et généraux, ils ne pourront jamais couvrir toutes les situations auxquelles un soldat peut faire face sur le terrain, ni donner une ligne directrice spécifique d'action pour chacune d'entre elles. Ensuite, appliquer ces principes demande de savoir les interpréter en fonction d'une situation donnée, pour ensuite décider d'une séquence d'actions satisfaisante. Cela demande de l'intuition (Sparrow 2007, Sharkey 2009), du recul et une bonne connaissance du contexte politique, économique

et militaire de la mission. Cela est déjà très difficilement réalisable par un soldat humain, cela l'est encore davantage pour un agent artificiel, en raison d'un manque d'intuition et de compréhension de l'environnement dans lequel il évolue.

Le questionnement s'articule autour de deux interrogations. La première question est celle de savoir s'il est possible de programmer les règles du droit international humanitaire. La deuxième question, déjà partiellement abordée, est celle de savoir qui est responsable lors de l'application de la programmation.

La généralité des principes de droit international humanitaire, source de complexité

La capacité à programmer le respect du droit international humanitaire est l'objet de vives polémiques. Les auteurs défendant la thèse de la possible programmation des principes du droit international humanitaire utilisent une série d'arguments visant à démontrer les avantages de la machine par rapport à l'être humain. Une première série d'arguments tiendrait à une supériorité intrinsèque des machines. Celles-ci auraient des réactions plus rapides en raison de leurs capteurs, d'une meilleure intégration des informations et de l'absence de fatigue et de sentiment. Une deuxième série d'arguments liée aux premiers milite en faveur d'une utilisation mieux contrôlée de la force. Les capteurs permettraient une meilleure appréciation des cibles et l'absence de sentiment et d'instinct de préservation conduiraient à un usage limité de la force car la machine ne construit pas des « scénarios » mais se contente de la réalité. En conclusion les machines conduiraient à une diminution des dommages collatéraux et des attaques erronées. À cette fin, Arkin (Arkin 2007) propose une architecture robotique qui se fonde sur la théorie de la guerre juste et qui a pour objectif de faire obéir le système à certains principes du droit international humanitaire et aux règles d'engagement spécifiques à la mission. Elle intègre plusieurs composants (*ethical governor, ethical behavioral control*, etc.) destinés à analyser les situations dans lesquelles se situe le système et à calculer une réponse adaptée et respectant un certain nombre de contraintes déterminées a priori. Cependant, l'approche proposée par Arkin nécessite des prérequis techniques (par exemple une bonne compréhension de la situation) trop importants pour être mis en œuvre dans des machines à l'heure actuelle. De plus, les grands principes structurants du droit international humanitaire – distinction, nécessité, proportionnalité, humanité – présentent des difficultés intrinsèques qui font obstacle à leur implémentation.

Le principe d'humanité, une relative clarté

Le principe d'humanité est lié à l'interdiction des maux superflus entendus comme les blessures, les destructions qui ne sont pas nécessaires pour atteindre le but recherché⁵. Si, comme toutes les nouvelles armes, les systèmes robotisés ne seront licites que s'ils ne causent pas de maux superflus, leur utilisation peut, cependant, se révéler en contradiction avec le principe d'humanité (Blank, 2012). Leurs avantages réels ou supposés - plus grande précision et meilleure utilisation de la force...- risquent de conduire paradoxalement à une utilisation plus permissive en suivant l'idée qu'ils seraient moins « nocifs » dans un environnement urbain ou semi-urbain par exemple. Ceci peut entraîner un accroissement corrélatif des victimes et des risques d'erreurs. Aussi, est-il possible de se demander si le droit international impose d'inclure dans la réflexion la prévisibilité ou l'imprévisibilité des dommages ou plus largement des implications pratiques des nouvelles armes (ce qui revient à réfléchir à leur intégration dans les méthodes de combats). Certes, l'article 36 du Protocole additionnel I impose d'examiner la conformité des armes nouvelles au droit international. Mais, en allant plus loin, sans être formellement interdites, la nature des dommages potentiels ou les méthodes de combats découlant de ces nouvelles armes pourraient-elles ou devraient-elles être prises en compte dans la réflexion ? Par exemple, la violation probable de droits de l'homme (droit à la vie, droit à la vie privée...) par l'utilisation de ces systèmes doit-elle conditionner leur interdiction totale ou partielle ? Mener une telle réflexion conduirait à imposer des obligations aux Etats qui iraient au-delà du droit positif. En conséquence, sauf si le concepteur programmait délibérément la machine pour causer des maux superflus, le principe d'humanité ne pose pas de véritable difficulté de programmation. Il sert de limite à ce que la machine peut faire.

Les principes de distinction et de nécessité, une implémentation ardue

Les difficultés relatives à la programmation se révèlent avec le principe de nécessité et le principe de distinction. Le principe de nécessité autorise le recours à la force nécessaire pour atteindre un objectif militaire et le principe de distinction oblige à distinguer les objectifs pouvant être attaqués de ceux qui ne le peuvent pas⁶. Son objet est de distinguer les objectifs militaires, qui peuvent être attaqués, des biens et populations civiles qui ne doivent, à l'inverse, faire l'objet d'aucune attaque volontaire⁷. Le principe de distinction impose ainsi aux parties au conflit de distinguer lors de leurs opérations les combattants et les biens militaires, qui peuvent être ciblés, des non-combattants et des biens civils, qui bénéficient d'une immunité contre les attaques. La détermination de la licéité de la cible, ou plus

⁵ Voir les articles 35 du Protocole additionnel I et 8 (2) (b) (xx) du Statut de Rome.

⁶ Voir les articles 48 et 51 du Protocole additionnel I.

⁷ Articles 48 et 52 du Protocole additionnel I. L'article 8 (2) (b) (i) et (ii) du Statut de Rome érige le non respect de cette distinction en crime de guerre.

exactement le caractère militaire de l'objectif, fait cependant difficulté dans les conflits armés contemporains. Selon le droit positif, les biens constituent des objectifs militaires soit en raison de leur nature soit parce qu'ils contribuent à l'action militaire d'une partie et que leur destruction offrirait un avantage militaire précis. Par exemple, une école utilisée par les forces adverses comme quartier général. De même, en vertu de ce principe, seuls les individus sans immunité constituent des objectifs militaires. Or, les catégories d'individus pouvant être directement ciblés sont en nombres réduits : les combattants (membres des forces armées conventionnelles) ou les « combattants illégaux⁸ » (Finaud 2006) que sont les mercenaires, les espions et les civils participant directement aux hostilités que ce soit au sein d'un groupe armé ou individuellement. C'est cette dernière catégorie qui est au centre des préoccupations des militaires et de la doctrine. La difficulté provient de l'incertitude entourant la notion de participation directe aux hostilités. Ni les Conventions de Genève ni leurs Protocoles additionnels ne précisent en effet quelle conduite constitue une participation directe aux hostilités. Dans son guide interprétatif relatif à la participation directe aux hostilités, le CICR a aussi distingué deux hypothèses dans lesquelles un civil perd son immunité contre les attaques (Melzer 2010).

La première hypothèse, qui n'est pas véritablement pertinente pour l'étude des robots, est celle des civils participants directement aux hostilités. Ces derniers perdent leur immunité contre les attaques dès lors qu'ils commettent, avec la volonté d'avantager une partie au détriment de l'autre, un acte hostile ayant un lien de causalité avec le dommage subi par l'une des parties. La distinction avec l'hypothèse suivante tient dans les caractéristiques de cette participation : spontanée, sporadique et non organisée.

La seconde hypothèse résulte de la civilianisation des conflits armés, un individu perd son immunité dès lors qu'il est membre d'un groupe armé. Un groupe armé est une structure hiérarchisée capable de mener des actions militaires. Cette appartenance est toutefois insuffisante. Pour perdre son immunité, l'individu devra, au sein du groupe, tenir une fonction continue de combat. Cette dernière condition est ouverte à de multiples interprétations plus ou moins restrictives. Car ce qui est en jeu, c'est de savoir si sont inclus dans cette fonction de combat les leaders politiques ou religieux d'un groupe armé.

Ces définitions montrent la difficulté de programmer les principes centraux du droit international humanitaire. En effet, que ce soit le principe de nécessité ou de distinction, un jugement *in concreto* est nécessaire. Or, il n'existe pas de normes ou de standards permettant de savoir comment implémenter des principes juridiques abstraits dans un système robotique pour les appliquer de façon autonome. Par essence les principes du droit international humanitaire et leur interprétation sont subjectifs et requièrent de prendre en compte un contexte trop complexe (politique, militaire, etc.) pour qu'ils soient codés et appliqués de façon fiable. Pour ce qui est de la programmation du principe de distinction, il est

⁸ Catégorie non juridique mais pédagogique.

théoriquement possible, selon certains, de programmer des listes fixes des cibles légitimes, par exemple des catégories de personnes ou d'armes. Néanmoins, techniquement la machine ne possède pas de moyen pour analyser visuellement son environnement et identifier de façon fiable les objets qui le composent. Les techniques actuelles sont très coûteuses et assez peu fiables compte tenu du bruit⁹ et de la multitude d'objets composant un environnement ouvert (par exemple, des bâtiments, des gravats ou des déchets en mouvement pour un environnement urbain). Qui plus est, des aléas environnementaux peuvent amoindrir les résultats (pluie, obscurité, etc.) Si l'on tient compte de la difficulté pour l'opérateur humain de percevoir clairement l'environnement physique du système, ce même via une caméra, on peut considérer qu'il est actuellement impossible d'appliquer correctement le principe de distinction sur la base d'indices visuels. Cela répond par là même à la question de savoir si la machine peut être programmée pour adopter un raisonnement inductif sur les caractéristiques des cibles légitimes hors de toute liste. En effet, comment une machine pourra-t-elle différencier un non-combattant, un combattant ou un civil participant directement aux hostilités ? Il semble que la présence d'un opérateur reste nécessaire, ce qui repose le problème du partage d'autorité.

Le principe de proportionnalité ou la complexité d'implémenter un jugement relatif

Les mêmes difficultés soulevées par les principes de nécessité et de distinction apparaissent pour le principe de proportionnalité. Celui-ci oblige à s'abstenir de lancer une attaque dont on peut attendre qu'elle cause incidemment des pertes en vies humaines dans la population civile, des blessures aux personnes civiles, des dommages aux biens à caractère civil, ou une combinaison de ces pertes et dommages, qui seraient excessifs par rapport à l'avantage militaire concret et direct attendu. Le principe de proportionnalité n'interdit donc pas tout dommage mais seulement les dommages excessifs faits aux civils par rapport à l'avantage militaire attendu. L'avantage militaire se définit comme l'avantage attendu de l'action. Il n'est pas nécessaire qu'il ait un rapport temporel ou géographique avec l'objet de l'attaque mais il doit être précis et réel. En effet, l'avantage militaire peut être interprété comme devant être défini soit par rapport à l'ensemble d'une stratégie de guerre et non en se limitant aux gains tactiques (Sassoli, Bouvier et Quintin 2012, p. 6), soit par rapport au mal à éviter (Parks 2010). Or, il ne s'agit pas d'une analyse quantitative, c'est-à-dire au regard du nombre de morts ou de blessés, mais d'une question de jugement.

⁹ Tout signal ou stimulus indésirable, comme du grain sur une image ou des grésillements de fond lors d'une communication audio.

La difficulté de l'analyse de proportionnalité réside dans la comparaison des dommages avec l'avantage militaire attendu, qui est l'élément central de l'évaluation de proportionnalité. Pour que la machine puisse respecter le droit international humanitaire, il faudrait qu'elle soit capable de mesurer l'équilibre entre l'avantage militaire et les dommages pour déterminer si en l'espèce ceux-ci seraient excessifs ; autrement dit, la machine devrait pouvoir s'interdire ou arrêter une attaque licite, en première analyse, car son résultat serait excessif.

L'implémentation, vecteur d'une pluralité de responsables

La deuxième question qui doit être abordée est celle de la responsabilité en cas de violation du droit international humanitaire : qui sera responsable dans cette hypothèse ? Sans reprendre les développements du paragraphe précédent, il convient de centrer l'analyse sur la responsabilité du concepteur. L'hypothèse selon laquelle la machine a été programmée sciemment pour violer le droit international humanitaire ne sera pas abordée car il est peu probable qu'une machine soit conçue pour commettre des crimes de guerre. Cependant, la vente d'une machine qui ne respecterait pas les prescriptions du droit international humanitaire pose la question de savoir qui du concepteur, de ceux qui l'ont testée, certifiée ou de l'utilisateur, sera responsable. Même si l'utilisateur final ne peut s'exonérer de sa responsabilité en arguant du fait qu'il a utilisé la machine selon les spécifications, la responsabilité de ceux qui ont contribué à la mise au point de la machine pourra être recherchée. Toutefois, la probable absence d'intention dans la commission de ces violations écarte la qualification de crime de guerre. Néanmoins, le fait qu'un tel système soit construit et déployé en sachant qu'il ne respecte pas le droit international humanitaire, notamment le principe de distinction, équivaut à accepter la réalisation de crimes de guerre. En cas de déploiement, la décision de recourir à la force ne devrait donc pas être celle des systèmes mais relever de l'opérateur.

La question de l'éthique

Si l'implémentation des règles juridiques régissant les conflits s'avère complexe, qu'en est-il de la fonction « éthique », autre aspect important susceptible de permettre une régulation du comportement d'un système autonome ?

Comme il a déjà été noté, l'éthique intervient quand apparaissent des dilemmes, quand des lois, des règles morales, entrent en conflit, ou encore lorsqu'il n'y a aucune règle pour nous guider.. Or, il est impossible de prévoir toutes les situations complexes qui vont se présenter, et les éventuels conflits de bien qui peuvent en résulter. La résolution d'un dilemme éthique se fait donc au cas par cas, en pesant les circonstances, les arguments en jeu, les règles juridiques concernées, les enjeux militaires, humains, etc.

Une machine peut-elle résoudre de tels conflits ? Pour Dominique Lambert (Lambert 2011), la réponse est non : la complexité du contexte, la subtilité infinie des détails, suggèrent qu'un programme informatique ne peut se substituer à un décideur. Pour approfondir la question, on peut en aborder une autre : que fait l'humain que ne peut faire la machine ? Cette question est évidemment trop complexe pour être développée ici, mais on peut évoquer quelques éléments de réflexion. En premier lieu, l'humain possède des émotions, qui interviennent beaucoup dans ses processus de décision. Il s'appuie dessus pour trancher. Implémenter des « émotions » dans des machines est une voie suivie par certains (Arkin 2007). En second lieu, l'humain possède de l'intuition, qui lui permet de « sentir » les situations au-delà de la rationalité, et prendre des décisions même en l'absence d'élément déterminant. En troisième lieu, l'humain possède de la créativité, qui lui permet d'inventer des solutions complètement innovantes. Citons ici un exemple rapporté par (Pichevin(2) 2012). « Dans le cadre d'une mission de rétablissement de la paix dans un pays en guerre civile, un détachement de quelques hommes avait installé son campement dans une zone ravagée par les combats et tous les troubles qui en découlent – famine notamment. Un jour, des civils hommes, femmes, enfants, vieillards, se sont avancés vers le campement. Ils criaient leur faim, certains étaient armés de machettes. Les ressources alimentaires du campement étaient calibrées pour les soldats, mais n'étaient pas prévues pour assurer le ravitaillement de cette foule. Que faire ? Une multitude de paramètres entraient en compte ici : certains étaient juridiques et moraux (ne pas tirer sur des civils), d'autres moraux (la compassion pour une population affamée), d'autres strictement militaires (accomplir la mission militaire de pacification). Le chef de détachement a décidé de recourir à l'intimidation, et fait poster ses soldats autour du camp, l'arme au poing. La foule a continué d'approcher. Le chef de détachement a commandé de charger les armes. La foule a fait demi-tour ». Cette description montre comment l'humain appréhende les problèmes éthiques. En l'état actuel des choses, une machine ne possède aucune de ces facultés, et ne saurait donc prendre de telles décisions. Un problème demeure néanmoins : dans quelle mesure ces facultés permettent-elles de prendre de « bonnes » décisions éthiques ? Que l'on envisage, par exemple, la question des émotions. Celles-ci sont ambivalentes : elles peuvent faire éprouver de la compassion – mais aussi de la haine.

Conclusion

L'intensification du recours aux systèmes sans équipage embarqué, qui a pour conséquence la robotisation du champ de bataille, laisse craindre que celui-ci voie de plus en plus de robots, plus ou moins indépendants, et de moins en moins d'êtres humains. Ceci pose inévitablement la question des places respectives de l'homme et de la technologie dans la conduite de la guerre. Pourtant, il est important de relativiser la nouveauté du phénomène car l'évolution technologique a historiquement toujours rythmée la guerre, même si, il est vrai, la volonté de cette nouvelle phase est bien celle de dépasser les limites de l'homme et par là même de conduire à son effacement. Cette ultime étape de l'autonomisation totale du robot n'est pour l'heure que de l'ordre de l'imaginaire et l'homme demeure bien présent, intellectuellement du moins, sur le champ de bataille. La révolution n'existe donc pas, du moins pour l'instant, l'évolution est celle de la présence d'un nouvel outil au service des armées.

Références

ARKIN, Ronald C., *Governing Lethal Behaviour: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture*, Georgia Institute of Technology, 2007.

BLANK, Laurie R., « After Top Gun : How drone strikes impact the law of war », *University of Pennsylvania Journal of International Law*, vol. 33, 2012, pp. 675-718

BOELLA, Guido, « Obligation and cooperation – Two sides of social rationality », in HEXMOOR, Henry, CASTELFRANCHI, Cristiano, and FALCONE, Rino (eds.), *Agent Autonomy*, Kluwer Academic Publishers, 2003, pp. 75-102.

BRADSHAW, Jeffrey M., SIERHUIS, Maarten, ACQUISTI, Alessandro, FELTOVICH, Paul, HOFFMAN, Robert, JEFFERS, Renia, PRESCOTT, Debbie, SURI, Niranjan, USZOK, Andrezej, and VAN HOOFF, Ron, « Adjustable autonomy and human-agent teamwork in practice : An interim report on space applications », in HEXMOOR, Henry, CASTELFRANCHI, Cristiano, and FALCONE, Rino (eds.), *Agent Autonomy*, Kluwers Academic Publishers, 2003, pp. 243-280.

BRAYNOV, Sviatoslav, and HEXMOOR, Henry, « Quantifying reactive autonomy in multiagent interaction », in HEXMOOR, Henry, CASTELFRANCHI, Cristiano, and FALCONE, Rino (eds.), *Agent Autonomy*, Kluwer Academic Publishers, 2003, pp. 55-73.

CASTELFRANCHI, Cristiano, and FALCONE, Rino, « From automaticity to autonomy: the frontier of artificial agents », in HEXMOOR, Henry, CASTELFRANCHI, Cristiano, and FALCONE, Rino (eds.), *Agent Autonomy*, Kluwer 2003, pp. 103-136.

CLEMENT, E., C. DEMONQUE, L. HANSEN-LOVE, P. KAHN, « Responsabilité », *La philosophie de A à Z*, Hatier, 2000

COOKE, Nancy J., and CHADWICK, Roger A., « Lessons learned from human-robotic interactions on the ground and in the air », in BARNES, Michael, and JENTSCH, Florian (eds.), *Human-Robot Interaction in Future Military Operations*, 2010, pp. 355-372.

CUMMINGS, Mary (Missy), « Automation and accountability in decision support system interface design », *Journal of Technology Studies*, vol 32:1, 2006

DAUPS, Thierry, “Point de vue juridique sur la mise en cause de l’autorité responsable au regard de l’autonomie et de la prévisibilité robotique”, novembre 2012.

DEPARTMENT OF DEFENSE, DEFENSE SCIENCE BOARD, “The role of autonomy in DoD systems”, July 2012 (www.fas.org/irp/agency/dod/dsb/autonomy.pdf)

DEPARTMENT OF DEFENSE, Directive Nr 3000.09 “Autonomy in Weapon Systems”, novembre 2012

DINSTEIN, Yoram, « Distinction and Loss of Civilian Protection in International Armed Conflicts », in *International Law and Military Operations*, CARSTEN, Michael D. (ed.), Rhode Island, Naval War College, 2008, pp. 183-198

FINAUD, Marc, « L'abus de la notion de "combattant illégal" : une atteinte au droit international humanitaire », *RGDIP*, 2006, pp. 862-889

GROS, Florian, TESSIER, Catherine, and PICHEVIN, Thierry, « Ethics and Authority Sharing for Autonomous Armed Robots », ECAI'12 Workshop on "Rights and Duties of Autonomous Agents (RDA2)", Montpellier, France, August 2012.

HALLEVY, Gabriel, « The Criminal Liability of Artificial Intelligence Entities from science fiction to legal social control », *Akron Intellectual Property Journal*, vol. 4, 2010, pp. 171-201

HANCOCK, Peter A., and SCALLEN, Stephen F., « Allocating functions in human-machine systems », in HOFFMAN, Robert R., SHERRICK, Michael F., and WARM, Joel S. (eds.), *Viewing Psychology as a Whole*, Washington, D.C. : American Psychological Association, 1998, pp. 509-540.

HEXMOOR, Henry, CASTELFRANCHI, Cristiano, and FALCONE, Rino, « A prospectus on agent autonomy », in HEXMOOR, Henry, CASTELFRANCHI, Cristiano, and FALCONE, Rino (eds.), *Agent Autonomy*, Kluwer Academic Publishers, 2003, pp. 1-10.

HEYNS Christof, "Rapport du Rapporteur spécial sur les exécutions extrajudiciaires, sommaires ou arbitraires", A/HRC/23/47, 9 avril 2013

HUANG, Hui-Min, PAVEK, Kerry, NOVAK, Brian, ALBUS, James, and MESSIN, Elena, « A framework for autonomy levels for unmanned systems ALFUS », in *AUVSI's Unmanned Systems North America 2005*, Baltimore, MD, USA, 2005.

HUMAN RIGHTS WATCH, "Losing humanity – The case against killer robots", November 2012, (http://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf).

JONES, Jeff, "Autonomy Levels for Unmanned Systems (ALFUS) Framework", 2008, (<http://sstc-online.org/2008/pdfs/JBJ2079.pdf>)

KANT, Emmanuel, *Fondements de la métaphysique des mœurs*, 1785. Vrin, 2004.

LAMBERT, Dominique, « Robots autonomes », in R. Doaré et H. Hude (dir), *Les robots au cœur du champ de bataille*, Economica, 2011.

LIN, Patrick, BEKEY, George, and ABNEY, Keith, *Autonomous military robotics: Risk, ethics, and design*, California Polytechnic State University, 2008.

LUCK, Michael, D'INVERNO, Mark, and MUNROE, Steve, « Autonomy: Variable and generative », in HEXMOOR, Henry, CASTELFRANCHI, Cristiano, and FALCONE, Rino (eds.), *Agent Autonomy*, Kluwer Academic Publishers, 2003, pp. 11-28.

MELZER, Nils, *Guide interprétatif sur la notion de participation directe aux hostilités en droit international humanitaire*, Genève, CICR, 2010, 88 p.

MERCIER, Stéphane, TESSIER, Catherine,, and DEHAIS, Frédéric, « Détection et résolution de conflits d'autorité dans un système homme-robot », *Revue d'Intelligence Artificielle*, numéro spécial « Droits et Devoirs d'Agents Autonomes », RSTI série RIA Volume 24 - 3/2010, mai-juin 2010.

MURPHY, Robin R., and BURKE Jennifer L., « The safe human-robot ratio », in BARNES, Michael, and JENTSCH, Florian (eds.), *Human-Robot Interactions in Future Military Operations*, Ashgate, 2010, pp. 22-24.

NAVAL RESEARCH LABORATORY, "A timeline of NRL's autonomous system research", 2011, (http://www.nrl.navy.mil/lasr/documents/Autonomous_timeline.pdf).

NEUBERG, M., « Responsabilité », in M. Canto-Sperber (dir.), *Dictionnaire d'éthique et de philosophie morale*, Puf, 2004.

PARKS, W. Hays, « Part IX of the ICRC "Direct Participation in Hostilities" Study : No Mandate, No Expertise, and Legally Incorrect », *International law and Politics*, vol. 42, 2010, pp. 769-830

PICHEVIN, Thierry, *Tourbillonnement éthique d'un océanographe*, Thèse de doctorat de l'université Paris XI, 2012.

PICHEVIN(2), Thierry, « Drones armés et éthique », in D. Danet, J.-P. Hanon, G. de Boisboissel (dir.), *La guerre robotisée*, Economica, 2012

PIZZIOL, Sergio, TESSIER, Catherine, and DEHAIS, Frédéric, « What's the heck is it doing? Better understanding Human-Machine conflicts through models », ECAP'12 Workshop on "Rights and Duties of Autonomous Agents (RDA2)", Montpellier, France, August 2012.

POMES, Eric, "La responsabilité des robots militaires", Projet de réflexion, 2012.

REGIS, Nicolas, DEHAIS, Frédéric, TESSIER, Catherine, and GAGNON, Jean-François, « Ocular metrics for detecting attentional tunnelling », Human Factors and Ergonomics Society Europe Chapter annual meeting, Toulouse, France, October 2012.

SARTER, Nadine, WOODS, David, and BILLINGS, Charlie, « Automation surprises », in *Handbook of Human Factors and Ergonomics*, 2nd edition, Wiley, 1997

SASSOLI, Marco, BOUVIER, Antoine A., QUINTIN, Anne, *Un droit dans la guerre ?*, vol. 1, 2^e édition, Genève, CICR, Référence, 2012

SCHRECKENGHOST, Debra, RYAN, Daniel, THRONESBERY, Carroll, BONASSO, Peter, and POIROT, Daniel, « Intelligent control of life support systems for space habitat », in *Proceedings of the AAAI-IAAI Conference*, Madison, WI, USA, 1998.

SHARKEY, Noel, « Death strikes from the sky : the calculus of proportionality », *Technology and Society Magazine*, IEEE, vol. XXVIII, n° 1, 2009, pp. 16-19.

SIMON, Herbert A., « Invariants of human behaviour », *Annual Review of Psychology*, n°41, 1990, pp. 1-19.

SPARROW, Robert, « Killer Robots », in *Journal of Applied Philosophy*, vol. XXIV, n° 1, 2007, pp. 62-77.

TESSIER, Catherine, and DEHAIS, Frédéric, « Authority management and conflict solving in human-machine systems », *AerospaceLab, The Onera Journal*, Vol. 4, 2012 (<http://www.aerospace-lab-journal.org/al4/authority-management-and-conflict-solving>)

TRUSZKOWSKI, Walt, HALLOCK, Lou, ROUFF, Christopher, KARLIN, Jay, RASH, James, HINCHEY, Michael G., and STERRITT Roy, “Autonomous and autonomic systems with applications to NASA intelligent spacecraft operations and exploration systems”, Technical report V2.0, July 2009.

WALLACH, Wendell, “Robot minds and human ethics: the need for a comprehensive model of moral decision making”, *Ethics of Information technology*, vol. 12, 2010, pp. 243-250

ZIEBA, Stéphane, POLET, Philippe, and VANDERHAEGEN, Frédéric, « Using adjustable autonomy and human-machine cooperation to make a human-machine system resilient – Application to a ground robotic system, *Information Sciences*, n°181, 2011, pp. 379-397.

Le Procureur C. Thomas Lubanga Dyilo, “Décision sur la confirmation des charges”, aff. n° ICC-01/04-01/06, 29 janvier 2007

Le Procureur c. Germain Katanga et Mathieu Ngudjolo Chui, “Décision relative à la confirmation des charges”, aff. n° ICC - 01/04 - 01/07, 30 septembre 2008

Le Procureur c. Jean-Pierre Bemba Gombo, “Décision rendue en application des alinéas a) et b) de l'article 61 - 7 du Statut de Rome, relativement aux charges portées par le Procureur à l'encontre de Jean-Pierre Bemba Gombo”, aff. n° ICC - 01/05 - 01/08, 15/06/2009