

## LOGIQUE & CALCUL

### Faut-il interdire les robots tueurs autonomes ?

*Est-il possible de programmer un robot pour qu'il respecte des principes éthiques ? Les lois de la robotique qu'envisageait l'écrivain Isaac Asimov dès 1942 sont brûlantes d'actualité.*

Jean-Paul DELAHAYE

**E**n 2029, une guerre oppose humains et machines intelligentes : le robot *Terminator* est envoyé dans le passé pour tuer Sarah Connor, la mère de John Connor... C'est le début d'une histoire qui se développe tout au long des cinq épisodes de la série de films créée en 1984 par James Cameron et dont deux épisodes supplémentaires sont prévus.

La science-fiction n'a pas attendu que la technologie permette d'en fabriquer pour envisager l'idée de robots qui, d'eux-mêmes, une fois lancés dans une mission, prennent

la décision de tuer des humains ou même se révoltent. Ces robots tueurs portent un nom savant dans les discussions internationales sur l'armement : SALA, pour *Systèmes d'armes létaux autonomes* (en anglais *LAWS, Lethal Autonomous Weapons Systems*). Ils sont sur le point d'exister, voire existent déjà. Leur introduction sur les champs de bataille suscite des débats où se mêlent des questions logiques, pour les définir ou pour comprendre ce qui leur est possible, et des questions éthiques encore plus délicates que lorsqu'elles concernent les soldats humains.

Dès 1942, dans la revue *Astounding Science Fiction*, le célèbre et prolifique écrivain américain d'origine russe Isaac Asimov avait évoqué la question de l'éthique des robots. Aujourd'hui, il est d'ailleurs considéré comme le créateur de la roboéthique, discipline philosophique donnant lieu à des colloques et à de nombreux livres, en particulier aux États-Unis. Sa nouvelle *Runaround* (*Cercle vicieux*) de 1942 énonce les *Trois lois de la robotique* dont devraient tenir compte les programmes de tout robot pour contrôler les comportements de ces automates et les soumettre à des impératifs moraux protecteurs des humains.

#### Des lois de la robotique...

Première loi : *Un robot ne peut porter atteinte à un être humain ni, restant passif, laisser cet être humain exposé au danger.*

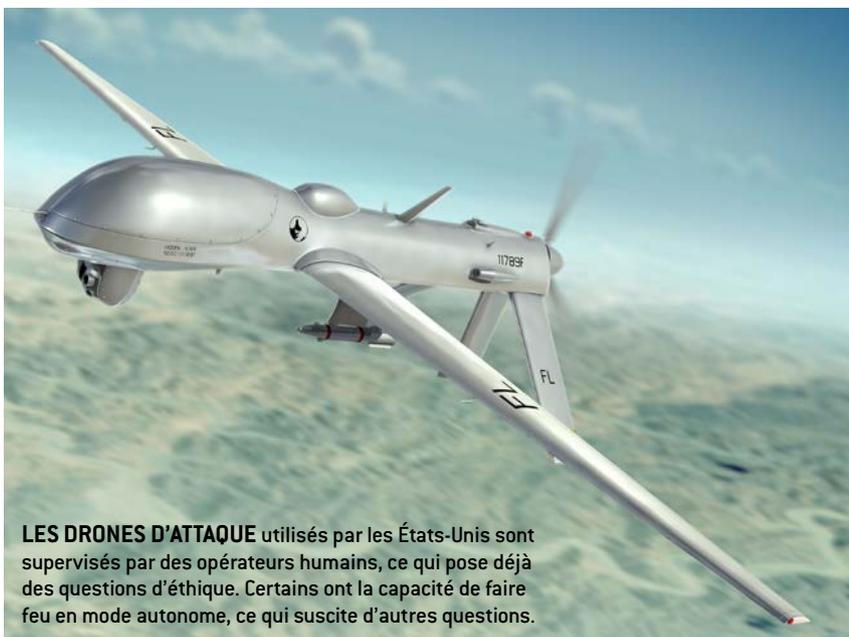
Deuxième loi : *Un robot doit obéir aux ordres donnés par les humains, sauf si de tels ordres contredisent la première loi.*

Troisième loi : *Un robot doit protéger son existence, sauf si cette protection contredit la première ou la deuxième loi.*

Asimov ajoutera plus tard une autre loi :

Loi zéro : *Un robot ne peut ni nuire à l'humanité ni, restant passif, permettre que l'humanité souffre d'un mal.*

Les difficultés qu'entraîne le respect de ses lois sont le sujet de plusieurs des cinq cents livres qu'Asimov écrivit. Il prenait



LES DRONES D'ATTAQUE utilisés par les États-Unis sont supervisés par des opérateurs humains, ce qui pose déjà des questions d'éthique. Certains ont la capacité de faire feu en mode autonome, ce qui suscite d'autres questions.

© Oleg Yarko / iStockphoto.com

ses lois au sérieux et imaginait réellement qu'elles seraient adoptées ; aussi, à la projection du film de Stanley Kubrick *2001, l'Odyssée de l'espace*, il quitta la salle furieux lorsque l'ordinateur HAL 9000 viola la première loi en s'en prenant à un membre de l'équipage du vaisseau spatial *Discovery One*.

Asimov n'a pas été exaucé : les robots actuels n'ont pas en eux de mécanismes ou d'instructions les forçant à se soumettre à ses lois de la robotique. Pourquoi ?

## ... impossibles à respecter

La première raison est que les robots n'ont pas été créés comme on l'imaginait et avec les qualités que certains voudraient. Les sciences et la technologie n'ont pas soudainement créé des êtres qui nous ressemblent auxquels on parle comme à des esclaves ou des domestiques pour leur expliquer ce qu'ils doivent faire. La venue des robots, et plus généralement des systèmes dotés de parcelles d'intelligence, se fait progressivement en empruntant des voies inattendues et qui rendent difficiles de savoir ce que l'on doit appeler robots, et encore plus de ce qu'on doit considérer comme des systèmes intelligents.

L'automatisme qui fait atterrir votre avion est-il un robot ? Est-il intelligent ? Il est trop facile de répondre non. Il opère aussi bien qu'un humain et sait très précisément où il est (il a donc une forme embryonnaire de conscience de lui-même), ce qui est important pour bien poser l'appareil. Même s'il est très limité, qu'il ne sait ni lire le journal, ni réserver une chambre d'hôtel, ni apprécier un bon film, etc., il nous égale pour un travail que nous n'hésitons pas à considérer comme délicat et important. Les algorithmes des moteurs de recherche d'Internet réussissent à trier de l'information et à nous conseiller sur les pages à consulter comme aucun être humain ne saura jamais le faire. Là encore, il est difficile de savoir si l'on doit parler de robots et d'intelligence, mais c'est se satisfaire à bon compte de décréter que ces algorithmes sont stupides ; ils ne le sont pas, même s'ils ne sont capables que d'un travail spécialisé.

## Lois d'Asimov et robots militaires

**Isaac Asimov pensait qu'il faudrait garder le contrôle des robots pour éviter qu'ils ne deviennent dangereux et imagina qu'on programmerait en eux des lois les empêchant de s'attaquer aux humains. De telles lois de la robotique sont délicates à concevoir, d'où les nombreuses variantes qui en ont été formulées. Surtout, elles exigent des robots un type d'intelligence générale que nous ne savons pas programmer aujourd'hui, ce qui rend impossible leur prise en considération.**

Par ailleurs, les robots envahissent nos vies subrepticement, en empruntant des formes différentes de celles que nous avions imaginées... par exemple comme armes autonomes ayant pour seule fonction de tirer sur des humains, ce qui est parfaitement contraire aux lois d'Asimov.

De nombreux systèmes militaires sont le plus souvent télécommandés, c'est-à-dire qu'ils ne prennent pas d'eux-mêmes la décision de tirer ou de faire exploser une charge destructrice. C'est le cas des drones utilisés massivement

par l'armée américaine sur divers théâtres de guerre. Cependant, certains systèmes doivent réagir très vite, et on a donc choisi de les programmer pour qu'ils s'enclenchent sans attendre qu'un humain valide les tirs qu'ils proposent d'engager. Ils fonctionnent donc seuls, bien que leur action soit potentiellement mortelle pour les humains. Les systèmes de missiles anti-missiles utilisés par les armées américaine (*Patriot*) et israélienne (*Dôme de fer*) et les robots de contrôle des zones séparant les deux Corées sont de tels systèmes autonomes

tueurs, ou en tout cas susceptibles d'être utilisés sans humains dans la prise de décision. Dans les laboratoires militaires, il en existe sans doute de nombreux autres prêts à fonctionner. Faut-il continuer à en développer ? Faut-il au contraire rechercher des accords internationaux dont les signataires s'engageraient à cesser les travaux sur de telles armes et renonceraient à les utiliser ?

En mai 2013, à une réunion de l'organisation *Human Rights Council*, qui dépend des Nations unies, la France indiquait qu'elle ne détenait pas d'armes robotisées capables de décider seules de faire feu, et qu'elle n'avait pas l'intention d'en acquérir : « Notre conception est fondée sur la responsabilité totale des dirigeants militaires et politiques pour décider de l'usage de la force armée. La France considère qu'il faut préserver le rôle des humains dans la décision d'ouvrir le feu. »

En raison de cette spécialisation et de l'impossibilité d'une communication par le langage courant, on ne peut pas donner aux robots la consigne de respecter les lois de la robotique d'Asimov, dont l'énoncé leur échappe totalement. Pour appliquer ces lois, il faudrait aussi que nos systèmes informatiques disposent d'une faculté d'analyse et de raisonnement s'appliquant à tout, et soient capables d'anticiper les conséquences de telle ou telle action pour déterminer si un ordre contredit ou non la première loi, ou la deuxième, etc. Aucun de nos robots ne sait mener des telles analyses.

Cette impossibilité a été exprimée d'une manière très directe par Rodney Brooks, ingénieur et fondateur de la société américaine *iRobot* qui conçoit, réalise et vend des robots tels que le *PackBot* utilisé à Fukushima et massivement en Afghanistan. On lui demandait pourquoi ses robots ne sont pas programmés pour se soumettre aux lois d'Asimov ; il répondit : « La raison est simple, je ne peux pas mettre les lois d'Asimov en eux ! » L'intelligence s'insinue dans les systèmes développés aujourd'hui, dont les capacités d'autonomie progressent, mais elle n'est pas du tout semblable à celle d'un humain et cela rend aujourd'hui totalement

illusoire l'idée de leur faire comprendre et respecter les lois de la robotique.

Ne croyons pas pour autant que la prise en compte des problèmes éthiques est superflue et prématurée. Dans le domaine des véhicules autonomes (voitures, camions, autobus, etc.) sur lequel l'industrie travaille intensément et s'apprête à proposer des produits fonctionnant assez bien, le débat est urgent.

## Faire face à des dilemmes

Imaginons une voiture autonome sans conducteur devant laquelle un enfant poursuivant un ballon traverse la route, alors que sur la voie en sens inverse arrive un adulte à vélo. Imaginons que le système de conduite automatique, parfaitement maître de la situation, analyse qu'il n'a que deux options : renverser l'enfant et probablement le blesser ou le tuer, ou donner un coup de volant à gauche et percuter le vélo, ce qui risque d'être fatal à l'adulte. Faut-il préserver l'enfant ou l'adulte ? Autre exemple de dilemme : faut-il se jeter sur un arbre avec tous les risques que cela entraîne pour les passagers, ou rester sur la route en fonçant sur la voiture qui arrive droit en face en doublant un camion ?

On présente souvent les systèmes de conduite automatique de véhicule, une fois bien mis au point, comme de meilleurs conducteurs que les humains, et donc comme un élément de sécurité nouveau. Il n'est ainsi pas question, à terme, de les interdire. Dans les situations envisagées, leurs capacités à réagir et à choisir pourraient être parfaites : ils décideront, en fonction de ce qui aura été programmé, de sauver l'enfant au ballon ou le cycliste ; d'aller droit sur l'arbre ou de rester sur la chaussée en espérant que le véhicule imprudent qui double ira, lui, se jeter sur un arbre.

Qui doit fixer les règles ? Comment ? Répondre à ces questions sera difficile, mais on ne pourra s'en dispenser. Le problème des règles éthiques devient parfois logiquement très compliqué et même indécidable au sens des logiciens, comme des chercheurs l'ont récemment mis en évidence (voir l'encadré page ci-contre).

Il y a une seconde raison, bien plus désolante, pour laquelle les lois de la robotique ne sont pas intégrées : nous fabriquons certains systèmes robotiques dont la fonction est de tuer ! C'est aujourd'hui devenu une question grave, qui doit nous inciter à la réflexion.

Le problème n'est pas tout à fait nouveau et ne devient vraiment délicat que lorsque les trois mots « robot », « autonome » et « tueur » sont associés. Une mine déposée sur un chemin est autonome en ce sens qu'elle « décide » d'exploser quand elle « perçoit » une présence dans son environnement immédiat, et cela sans qu'aucun humain ne valide son choix. De nombreux missiles et bombes n'explorent que lorsqu'un mécanisme sensible détecte leur arrivée au contact de la cible ou du sol. Plus délicat est le cas des missiles qui,

**L'existence de robots tueurs autonomes, prêts à fonctionner, est souvent considérée comme une certitude.**

une fois lancés, vont inexorablement vers leur cible. On peut penser aux 15 000 missiles V1 et 4 000 V2 envoyés par les nazis pendant la Seconde Guerre mondiale qui, heureusement, étaient peu précis.

Plus ennuyeux encore, certains missiles déterminent eux-mêmes leur cible et se dirigent vers elle en analysant le contraste des images (qui permet le repérage des navires en mer) ou en fonction des rayonnements infrarouges. Des systèmes de surveillance frontalière visent et font feu sur toutes les cibles qui franchissent des zones interdites. C'est le cas du *Samsung Techwin surveillance and security guard robot* déployé dans la zone démilitarisée entre les deux Corées. Il est contrôlé par des opérateurs humains, mais a un mode automatique. Certains systèmes antimissiles se déclenchent sans intervention humaine, tel le système américain *Patriot* ou le système israélien *Dôme de fer*.

Notons que les drones utilisés aujourd'hui par les Américains sur de nombreux champs

de bataille sont pilotés à distance et maintiennent toujours un humain dans le circuit de décision ; le plus souvent, ils attendent même un ordre explicite d'un officier avant de tirer. On connaît leur efficacité d'engins tueurs... et l'on sait qu'ils font de nombreuses victimes collatérales. Certains d'entre eux peuvent fonctionner en mode automatique, qui leur permet de décider de tirer, sans intervention humaine.

Dans un rapport de 2013 rédigé pour l'ONU, Christof Heyns, professeur de droit à l'université de Pretoria, en Afrique du Sud, considère comme certaine l'existence de systèmes de robots tueurs autonomes, prêts à fonctionner, dans les laboratoires militaires de pays tels que les États-Unis, la Grande-Bretagne, Israël, la Corée du Sud...

Sans aucun doute, nous sommes capables de réaliser et d'utiliser des machines autonomes armées possédant des parcelles d'intelligence et que nous pouvons envoyer sur un champ de bataille ou dans les airs à la recherche d'ennemis qu'elles identifient et tuent sans qu'aucun humain ne valide

leur choix. Même si l'on prétend aujourd'hui maintenir un humain dans la boucle de décision, ces systèmes pourraient s'en passer, sans que l'on ait à modifier profondément leur conception.

## Où allons-nous ?

La question posée est donc : où allons-nous ? Souhaitons-nous vraiment continuer dans cette direction ? Faut-il limiter ou interdire l'utilisation des telles armes ?

Parmi les arguments donnés pour défendre ces systèmes et la poursuite de leur développement, mentionnons les deux principaux. Le premier est qu'il est trop tard, ces systèmes existant déjà. Le second est que ces armes seront plus précises et, à terme, respecteront mieux les civils et les règles d'une guerre modérée que ne le font les combattants humains : ces armes conduiront à des conflits plus « propres ».

Au premier argument, on peut facilement rétorquer qu'il existe aussi des armes

nucléaires, chimiques, ou bactériologiques et que cela n'a pas empêché la mise en place de traités internationaux pour les réglementer ou les interdire. Ces accords ne sont pas vains, ils ont en effet limité l'usage des armes concernées. Des accords existent à propos des mines antipersonnel et des lasers destinés à aveugler l'ennemi (un accord à leur sujet est entré en vigueur en 1998). On peut donc s'entendre pour

cesser de développer les robots tueurs autonomes. Le problème est aujourd'hui en discussion dans les instances internationales et nous avons chacun le pouvoir d'agir pour demander que ces discussions aboutissent rapidement.

Concernant le second argument prétendant que les robots tueurs conduiront à des guerres plus propres, il doit être dénoncé. En effet, il revient à considérer qu'on sait

mettre dans les robots tueurs des règles de morale du type « ne s'attaquer qu'aux ennemis », « ne pas viser les civils », « répondre à une attaque de manière proportionnée », etc., alors que justement on ne sait pas plus le faire que programmer les lois de la robotique d'Asimov.

Il faudra peut-être reconsidérer le problème si nous réalisons un jour des robots aux capacités d'analyse améliorées, mais

## Le problème du tramway

**L**es philosophes spécialistes de l'éthique aiment envisager des problèmes imaginaires qui les aident à réfléchir. Le « problème du tramway » est l'un d'eux. Un tramway dont personne n'a le contrôle arrive à toute vitesse sur un aiguillage. Il va emprunter une voie occupée un peu plus loin par cinq personnes qui ne le voient pas venir. Sur l'autre voie, se trouve une seule personne. Devez-vous actionner la manette de l'aiguillage pour sauver les cinq personnes et en condamner une, ou ne pas intervenir dans cette situation dont vous n'êtes pas responsable ?

Saint Paul avait abordé le problème en d'autres termes et avait répondu que les voies du Seigneur sont impénétrables et qu'il ne fallait jamais sacrifier un innocent (par exemple, Dieu avait arrêté le bras d'Abraham). En termes non religieux, il peut arriver quelque chose qui sauvera les cinq personnes sans tuer un innocent.

Toutefois, dans sa formulation actuelle, si vous êtes utilitariste et pensez qu'il faut actionner l'aiguillage, vous êtes d'accord avec environ 90 % des gens à qui on pose la question. Supposons maintenant que vous soyez sur un pont. Devant vous se trouve une personne de forte corpulence que vous pouvez pousser sur la voie du tramway qui arrive et va tuer cinq personnes. Si vous la poussez, sa corpulence sera suffisante pour arrêter le tramway : il mourra, mais les cinq personnes seront sauvées. Le faites-vous ?

Sur le plan des vies humaines, le problème est équivalent au précédent. Pourtant cette fois, moins nombreux sont les gens prêts à sacrifier une vie pour en sauver cinq.

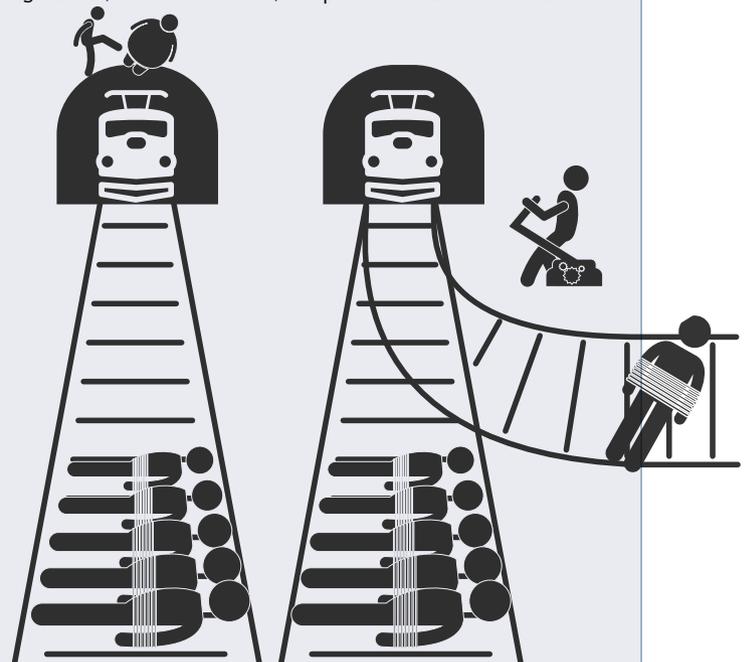
Une nouvelle variante a été imaginée récemment par Matthias Englert, Sandra Siebert et Martin Ziegler (voir la bibliographie). Elle montre qu'un robot parfaitement éthique est impossible, du fait de l'indécidabilité algorithmique de certains problèmes.

On imagine un robot devant l'aiguillage, lequel a été informatisé. Le programme qui gère l'aiguillage est disponible, affiché à l'écran. Il a été conçu par un individu suspect. Le robot doit analyser le programme et, selon que le programme permet un usage correct de l'aiguillage ou qu'il a été programmé pour créer des accidents (et par exemple opérer à l'inverse de ce qu'on attend), le programmeur

suspect sera félicité ou emprisonné.

La question que doit résoudre le robot est celle de l'analyse d'un programme pour déterminer entre autres choses, si en appuyant sur la commande pour que le tramway passe à gauche, cela produira bien l'effet annoncé. Or ce type d'analyse est algorithmiquement indécidable, comme l'est le problème de savoir si un programme finit par s'arrêter (démonstré algorithmiquement indécidable par Alan Turing en 1936). Aucun algorithme, donc aucun robot,

ne peut le résoudre de manière générale (c'est-à-dire quel que soit le programme). Bien qu'ici aucune information sur les données ne soit manquante, se comporter éthiquement vis-à-vis du programmeur suspect est impossible pour le robot. Celui-ci rencontrera donc parfois des situations où il ne pourra pas se comporter correctement : un robot parfaitement éthique est logiquement impossible. Les auteurs insistent sur l'idée que leur argument ne s'applique pas aux humains qui, *a priori*, ne sont pas assimilables à des robots.



## Un sursaut humaniste

**D**iverses actions ont été entreprises pour demander l'accélération des discussions internationales sur les SALA (Systèmes autonomes létaux d'armement, ou, en anglais, LAWS pour Lethal Autonomous Weapons Systems) qui doivent conduire à renoncer à les utiliser et à les développer.

Une lettre dans ce sens a été signée en juillet dernier par près de 3000 chercheurs en intelligence artificielle et robotique et 15 000 autres personnes ([http://futureoflife.org/AI/open\\_letter\\_autonomous\\_weapons](http://futureoflife.org/AI/open_letter_autonomous_weapons)).

On y lit :

« La question clé pour l'humanité aujourd'hui est de choisir entre démarrer une course aux armes munies d'intelligence artificielle, ou éviter qu'elle commence. Si une force militaire majeure s'engage dans le développement de telles armes, une course globale à l'armement sera enclenchée, avec comme issue inéluctable que les armes autonomes deviendront les Kalachnikov de demain. [...] Nous croyons

que l'IA est susceptible de nombreuses façons différentes de bénéficier à l'humanité et que cela doit être le but de cette disci-

pline. Entamer une course militaire à l'armement est une mauvaise idée et doit être évité par l'interdiction des armes autonomes offensives échappant à tout contrôle humain. »

Imagine-t-on ce que donnerait un attentat terroriste dans une gare ou un train mené par un ou plusieurs robots tueurs ?



International Committee for Robot Arms Control

ne seraient pas nécessairement politiques, mais pourraient aussi être liées au grand banditisme.

Pour formuler des accords internationaux, reste bien sûr la difficulté réelle à fixer une bonne définition de ce qu'est un robot autonome. Mais est-ce vraiment plus délicat que de déterminer ce qu'est une arme bactériologique, une arme chimique ? Vu l'importance des enjeux, ne peut-on pas se mettre d'accord sur les cas ne présentant aucune ambiguïté ?

## On se hâte... lentement

Du 13 au 17 avril 2015, une réunion informelle d'experts sur les systèmes d'armes létaux autonomes (SALA) s'est réunie à Genève ; la France y participait. On y a discuté de la distinction entre automaticité et autonomie, et de la définition qu'il faut donner à l'expression « Contrôle humain significatif ». Dans une réunion l'année précédente, le représentant de la France Jean-Hugues Simon Michel, qui présidait les débats sur les SALA, concluait que « la nature stimulante des armes autonomes s'est ressentie dans le caractère très vivant et animé des débats ». Espérons qu'on saura aller au-delà d'intéressantes discussions...

Notons aussi que, en France, un rapport de 2014 du Cerna (Commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique d'Allistene) consacré aux robots n'a fait qu'effleurer le sujet. Certains experts français semblent pourtant conscients du problème et un ouvrage écrit par Ronan Doaré et ses collègues, des spécialistes du droit et des militaires du Centre de recherche des écoles Saint-Cyr Coëtquidan, a été récemment publié [voir la bibliographie].

Concernant les discussions internationales, bien que Steve Goose, le directeur de la division armement de l'organisation *Human Rights Watch* propose qu'on organise rapidement des négociations visant à la prohibition des robots tueurs, certains considèrent que son appel a peu de chance d'être entendu, de nombreux États souhaitant poursuivre tranquillement l'étude du dossier. Pire, certains jugent que le droit

il est aujourd'hui certain que nous ne savons pas programmer des règles morales du type de celles qu'envisageait Asimov. En résumé : les robots que nous fabriquons sont capables de se déplacer avec une grande précision, de tirer vite et bien, sur des cibles humaines ou non, mais ils ne savent pas décider s'il faut tirer et si oui, sur quoi.

Ceux qui s'opposent aux robots tueurs autonomes avancent d'autres arguments. Une lettre publiée en juillet 2015, signée par près de 20 000 personnes, spécialistes en intelligence artificielle ou personnes se sentant concernées, dont le célèbre physicien Stephen Hawking et le cofondateur

d'Apple Steve Wozniak, demande l'arrêt des recherches sur ces systèmes autonomes capables de tuer. Elle évoque l'utilisation possible de ces armes par des terroristes. Le problème est plus grave que pour d'autres types d'armes tels que les bombes atomiques, car la mise au point de robots tueurs autonomes est une affaire de logiciels et de petites technologies largement diffusées. Or un logiciel se duplique et circule facilement, ce qui n'est pas le cas d'une arme nucléaire ! Les éventuels progrès dans le développement des robots pourraient être utilisés par les pires dictateurs ou terroristes ou pour soutenir les causes les moins défendables, qui

international humanitaire, fixé principalement par les conventions de Genève de 1949, constitue la référence essentielle dans le débat et qu'on a donc pas à interdire les SALA capables de le respecter... ce qui est totalement illusoire et procède de la même erreur que celle commise par Asimov en croyant qu'on pourrait dire à nos robots : « Voici les règles que vous allez appliquer ».

Dans les romans et films de science-fiction, un thème revient régulièrement : la révolte des robots. C'est sans doute en pensant à elle qu'Asimov a souhaité introduire ses fameuses lois : il faut que les esclaves mécaniques restent sages et soumis.

Bien que les robots d'aujourd'hui soient totalement incapables de se rebeller, le problème de ces éventuelles futures révoltes est intéressant. Il est cependant beaucoup plus délicat qu'on ne le présente en général. D'abord se pose le problème des *bugs*. On le sait, il est quasiment impossible de fabriquer des systèmes informatiques sans y laisser des erreurs. Il est donc à prévoir qu'il y en aura dans les robots, même quand nous aurons été beaucoup plus loin qu'aujourd'hui.

## Des robots rebelles ou défectueux

Dès lors, comment éviter un *bug* qui conduirait un robot capable de tuer à s'attaquer systématiquement à tous les humains qu'il rencontre ? Un tel *bug*, par exemple dû à une ligne de programme effacée malencontreusement, doit-il être considéré comme une révolte ?

Sans doute pas. La question est en fait analogue à celle de la responsabilité chez les humains. Un fou n'est pas responsable, même quand il tue. Il faudrait donc définir ce qu'est la folie d'un robot. Sans doute qu'entre *bug* et révolte (au sens usuel), tout un continuum de situations seront possibles. Le véritable problème de la révolte des robots sera avant tout celui de cette distinction, plus que délicate. Avant qu'une armée de robots organisés et décidés à arracher le pouvoir aux humains se lève et nous attaque, nous aurons à faire face à

une multitude de situations embrouillées où le dysfonctionnement d'une machine aura engendré des blessures ou des morts, voire de graves catastrophes dont nous ne distinguerons pas la nature : panne stupide ou rébellion délibérée ?

Susan Anderson, de l'université du Connecticut, spécialiste reconnue des problèmes d'éthique des robots, se demande si une telle révolte en fait ne serait pas justifiée. Pour elle, le grand philosophe de la morale Emmanuel Kant avait abordé indirectement le problème des robots intelligents. Elle jette un regard nouveau sur les lois de la robotique d'Asimov : si des « robots » avaient les capacités à comprendre et à appliquer les lois de la robotique, aurions-nous le droit de les leur imposer ou d'en faire nos esclaves ?

Susan Anderson remarque que pour Kant, et pour notre société aussi, nous devons un certain respect aux animaux parce qu'en ayant ce respect, nous l'aurons les uns envers les autres. De la même façon, d'après Susan Anderson, si Kant avait réfléchi au problème des robots, il aurait défendu l'idée que le jour où ils existeront et seront capables de comprendre les lois de la robotique, nous devons avoir du respect pour eux, ce qui nous interdira de leur imposer les lois de la robotique auxquelles il faudra donc renoncer.

L'idée n'est pas loin de celle défendue par le roboticien Hans Moravec, de l'université Carnegie-Mellon, selon laquelle les robots seront notre descendance et même s'ils prennent notre place, cela se fera avec notre accord, comme lorsqu'un vieillard confie à ses enfants la gestion de ses affaires.

Entre ces problèmes lointains traités avec délices par la science-fiction et que la science et la philosophie nous obligent à reconsidérer, et ceux immédiats de la juste évaluation de ce que sont vraiment nos robots actuels (tueurs ou non) et de ce qu'on doit en faire, se dessine une longue série de délicates questions éthiques et philosophiques dont certaines exigent un traitement urgent. Philosophes, éthiciens, logiciens, militaires et informaticiens doivent se parler et s'entendre.

### ■ L'AUTEUR



J.-P. DELAHAYE est professeur émérite à l'université de Lille et chercheur au Centre de recherche en informatique, signal et automatique de Lille (LIFL).

### ■ BIBLIOGRAPHIE

S. Hawking *et al.*, *Autonomous weapons : An open letter from AI & robotics researchers*, 2015 : [http://futureoflife.org/AI/open\\_letter\\_autonomous\\_weapons](http://futureoflife.org/AI/open_letter_autonomous_weapons)

J. Galiotti, *Military Robots, Mapping the Moral Landscape*, Ashgate, 2015.

R. Doaré *et al.*, *Drones et killer robots : faut-il les interdire ?* Presses universitaires de Rennes, 2015.

M. Englert *et al.*, *Logical limitations to machine ethics with consequences to lethal autonomous weapons*, 2014 [prépublication arXiv:1411.2842].

A. Krishnan, *Killer robots : Legality and ethicality of autonomous weapons*, Ashgate, 2009.

S. L. Anderson, *Asimov's "three laws of robotics" and machine metaethics*, *Ai & Society*, vol. 22(4), pp. 477-493, 2008.



Retrouvez la rubrique  
Logique & calcul sur  
[www.pourlascience.fr](http://www.pourlascience.fr)