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'explosent que lorsqu'un mécanisme sensible détecte leur arrivée au contact de la cible ou du sol. Plus délicat est le cas des missiles qui, 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.

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

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

Les 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 (démontré 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.