

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.

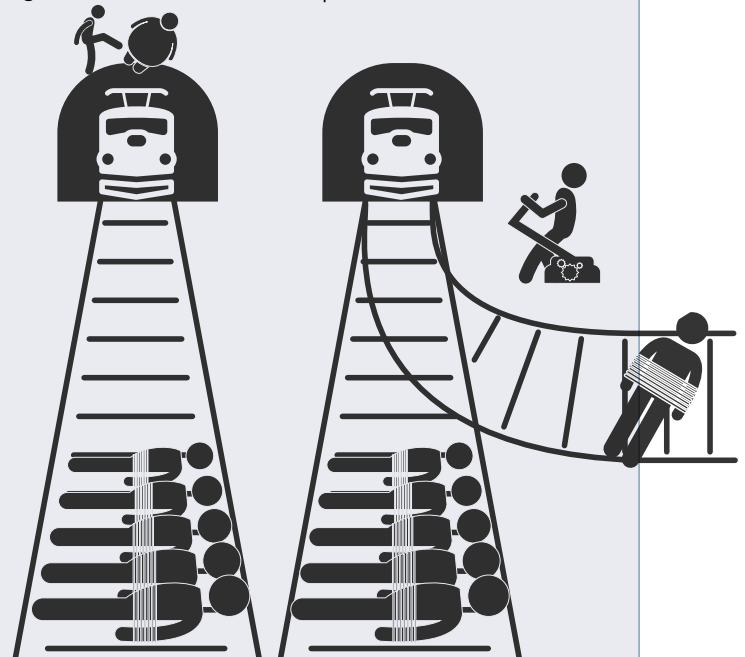
Une nouvelle variante a été imaginée récemment par Matthias Englert, Sandra Siebert et Martin Ziegler (voir la bibliographie). Elle montre qu'un robot parfaitement éthique est impossible, du fait de l'indécidabilité algorithmique de certains problèmes.

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 (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

Diverses 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).

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