

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.

