

itinéraire

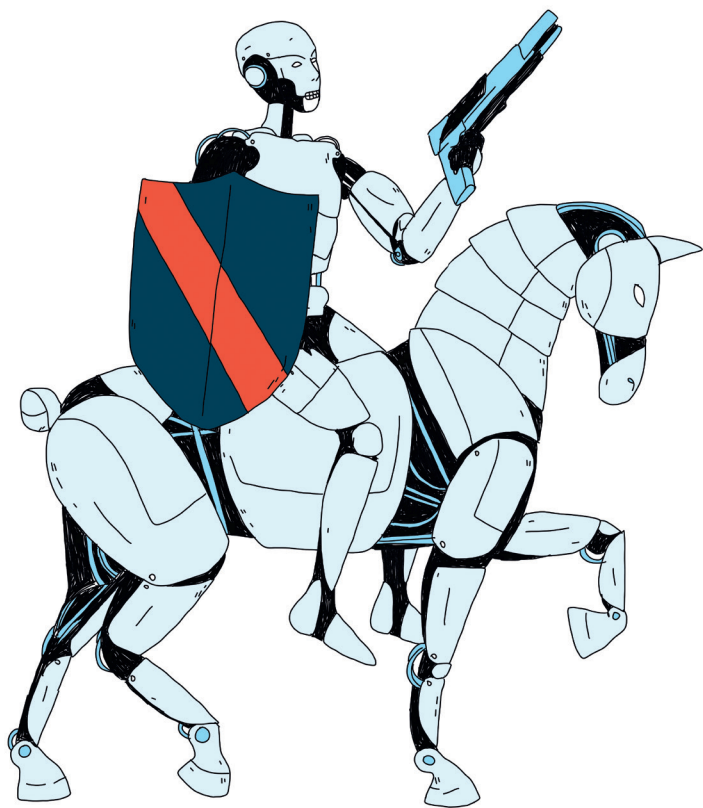
SI L'ON POUVAIT AUTREFOIS SE DEMANDER CE QUE LES ORDI-

d'un robot tueur

NATEURS NE SAVAIENT PAS ENCORE FAIRE, AUJOURD'HUI IL EST URGENT DE DÉFINIR CE QUE LES

m.-d.-n. ruffo

MACHINES NE DEVRAIENT PAS FAIRE. NOTAMMENT LES "SALA"...



[ESSAI LE POMMIER !]

itinéraire
d'un robot tueur

Marie-des-Neiges Ruffo de Calabre est docteur en philosophie de la Sorbonne et professeur de philosophie à l'université de Namur ainsi qu'à l'Institut catholique de Paris. Ses recherches portent sur les questions éthiques relatives à la sécurité, la défense et les nouvelles technologies, en ce compris l'intelligence artificielle. Récompensée d'un Prix scientifique de l'Institut des hautes études de défense nationale (IHEDN), elle intervient dans de nombreuses conférences internationales, notamment lors de la conférence annuelle de la CIOR (cadre OTAN) en 2018.

Visuel de couverture: Anna Wanda Gogusey

Couverture: Bianca Gumbrecht/Lunapark

Mise en page: Marina Smid

Préparation de copie: Valérie Poge

© Éditions Le Pommier/Humensis, 2018

Tous droits réservés

ISBN: 978-2-7465-1662-5

170 bis, boulevard du Montparnasse – 75014 Paris

www.editions-lepommier.fr

itinéraire d'un robot tueur

m.-d.-n. ruffo



[ESSAI LE POMMIER!]

À Joseph, toujours.

INTRODUCTION

Les robots ont-ils rendu l'humanité obsolète ? Les métiers les plus divers sont aujourd'hui menacés de disparaître, ou, tout du moins, risquent de ne plus être exercés par des humains. Que vous soyez médecin, caissière, voire prêtre bouddhiste¹, la robotique a prévu une solution de remplacement pour vous.

Si naguère la robotique venait compléter la chaîne d'assemblage, permettant ainsi à l'homme d'échapper progressivement à la dimension abrutissante de la division du travail, dont se moquait déjà Charlie Chaplin dans le film *Les Temps modernes*, aujourd'hui le robot investit le secteur tertiaire de l'économie. Dorénavant, de plus en plus de services à la personne seront exercés par des robots humanoïdes. Nao, Pepper, Asimo ou encore HRP-4² seront les nouveaux prénoms de vos infirmières, de vos vendeurs, de vos aides ménagères. Quant aux métiers dévolus aux gynoïdes, la version féminine des androïdes, ils témoignent de préjugés de genre bien ancrés : hôtesse d'accueil, mannequin ou prostituée – si le robot est révolutionnaire, le machisme a la dent dure.

Les robots sont-ils les nouveaux « casseurs de grèves », prêts à travailler pour bien moins cher que leurs homologues humains ? Vont-ils détruire l'humanité comme l'annonçait à la télévision le robot Sophia³ pour le plus grand embarras de son concepteur ? L'intelligence artificielle, abrégée I. A., va-t-elle dépasser l'intelligence humaine ? Pour se convaincre des capacités surhumaines de l'I. A., rappelons que le programme Deep Blue a battu le Russe Kasparov aux échecs et qu'AlphaGo a battu le Coréen Lee Sedol, l'un des meilleurs joueurs au monde, au jeu de go. À ce jour, le seul adversaire capable de battre AlphaGo n'est autre qu'un autre programme : AlphaGo Zero. *Zero* pour « zéro programmation » : cette nouvelle version a littéralement appris toute seule à jouer, sans humains pour la diriger une fois les règles du jeu connues.

Pour résister à la puissance des machines, devons-nous également effectuer notre propre « mise à jour » ? Devons-nous avoir recours à des techniques d'augmentation pour être autorisés à vivre dans ce nouveau monde ? L'humain de demain sera-t-il un cyborg ? C'est ce que suggère le courant transhumaniste. La pensée transhumaniste a deux facettes. Du côté face, elle dénigre ce qui fait le propre de l'humain, sa fragilité, sa mortalité, etc. Du côté pile, elle idolâtre l'évolution technologique, la puissance de calcul des ordinateurs, etc.

Cette possibilité que les robots deviennent « meilleurs » que les humains – plus efficaces, plus rapides, plus résistants – provoque une nouvelle blessure narcissique chez l'homme. Lui qui ne peut se définir qu'en se distinguant se retrouve cerné d'une part par sa proximité génétique (98,7 %) avec le chimpanzé et le bonobo, et d'autre part par cette intelligence artificielle qui semble plus raisonnable que lui et qui ne cesse

d'évoluer. Quelle est encore la spécificité de l'humain entre le machinal et l'animal ?

Nous nous sentons dépassés par ces technologies lorsque nous ne les comprenons pas. La peur de la technologie vient de cette ignorance des capacités réelles de la machine. Il est temps de se poser les vraies questions. Si le philosophe Hubert Dreyfus exposait dans son livre⁴ en 1972 « ce que les ordinateurs ne peuvent *pas encore* faire », à présent il est temps de nous demander : « Qu'est-ce que les ordinateurs ne devraient pas faire ? » Si les machines peuvent tout faire, que devrions-nous accepter qu'elles fassent ? À de rares exceptions près – les mines antipersonnel en sont un exemple – aucune technologie n'est mauvaise en elle-même. Tout dépend de l'utilisation qui en est faite.

Parmi les utilisations de la robotique les plus éthiquement problématiques à ce jour se trouvent les robots militaires. Le P.-D.G. de Tesla, Elon Musk, qu'on ne saurait taxer de technophobie, a mis en garde les Nations unies, avec d'autres experts, contre les SALA, « les systèmes d'armes létales autonomes »⁵. Plus connus sous le nom de « robots tueurs ». Pourquoi l'attention médiatique se concentre-t-elle sur l'utilisation militaire plutôt que sur les logiciels de trading automatisés, pourtant responsables d'un *flash krach* à mille milliards de dollars en 2010 ? Parce que dans une guerre, plus qu'ailleurs encore, les dimensions du respect de l'éthique, de la délégation de la décision à un programme informatique, de la place encore réservée à l'humain ne sont pas simplement des débats de mots mais des enjeux de vie ou de mort.

Le robot militaire constitue donc un cas limite focalisant l'attention sur les enjeux majeurs du remplacement de

l'humanité par des robots, dans ses tâches, ses décisions et ses devoirs. De façon générale, on peut dire que l'introduction des robots dans notre quotidien doit être évaluée dans un espace à trois dimensions : éthique, légal et pratique. Quant au robot, s'il désigne un système doté de capteurs, de processeurs et d'actionneurs, le mot signifie « esclave » et tire son origine de la pièce de théâtre *R. U. R.* (1920) du Tchèque Karel Čapek. Ce dernier affirmait : *« Je rejette avec horreur toute idée selon laquelle des engins de métal pourraient jamais remplacer les êtres humains et s'éveiller à des sentiments tels que la vie, l'amour ou la rébellion. Une prospective aussi sombre ne serait jamais qu'une surestimation du potentiel des machines ou encore une grave offense à la vie⁶. »* Mais Čapek écrivait cela en 1935. L'évolution technique a-t-elle rendu ce remplacement plausible aujourd'hui ?

Nous traduirons ce contexte à travers la métaphore d'une compétition entre l'homme et la machine. Le robot et l'humain se lancent un défi : le meilleur remplacera l'autre à la guerre. Pour savoir qui, de l'humain ou du robot, doit remplacer l'autre, un itinéraire avec trois lieux à traverser est établi : l'agora des philosophes, le tribunal des avocats et le champ de bataille des militaires. Qui sera considéré comme le meilleur combattant à l'issue de ce trajet et de ces joutes verbales ? Que le voyage commence...

1. La société japonaise Nissei propose d'utiliser le robot Pepper pour chanter les prières des cérémonies funéraires bouddhistes.
2. Nao, Pepper, Asimo, HRP-4 sont tous les quatre des robots humanoïdes. Initialement, Nao est une création française de la société Aldebaran Robotics, destiné à la recherche et à l'éducation, car il peut être programmable par l'utilisateur. L'entreprise a été rachetée par le japonais SoftBank Robotics. Le modèle Pepper, qui sert de robot-compagnon, a été produit par la même société. Asimo est une réalisation de Honda, il peut marcher seul et apporter un café dans un gobelet. HRP-4 est un robot d'assistance, il est destiné à aider les personnes âgées et à remplacer les travailleurs humains. Il est développé conjointement par l'Institut national des sciences et technologies industrielles avancées du Japon et Kawada Industries. À ne pas confondre avec le modèle plus connu HRP-4C, qui a une forme gynoïde.
3. Robot de conversation à visage féminin et au cerveau informatique visible, Sophia s'est vu attribuer la citoyenneté saoudienne.
4. *L'Intelligence artificielle: mythes et limites*, Flammarion, 1984. Titre original: *What computer still can't do*.
5. Le 21 août 2017, Elon Musk et d'autres experts ont signé une lettre ouverte aux Nations unies pour demander l'interdiction des robots tueurs.
6. Cité dans D. Ichbiah, *Robots, genèse d'un peuple artificiel*, Minerva, 2005.

SUR L'AGORA

Notre robot et notre humain entamèrent donc leur itinéraire et arrivèrent au premier lieu fixé pour leur compétition : l'agora des philosophes. L'humain, confiant, se planta au milieu de la place et cria à la cantonade :

— Ô sages ! dites-nous qui, du robot ou de moi-même, est le meilleur ? Celui que vous désignerez aura désormais le privilège de combattre au nom de la démocratie !

En réponse, un philosophe s'avança :

— Sais-tu bien, toi qui nous prends ainsi à partie en interrompant nos échanges, que nous ne pouvons désigner de meilleur que celui qui agira selon l'excellence ?

— Qu'appellez-vous l'excellence ? demanda le robot.

Relevant le pan de son manteau sur son épaule droite, le sage répond :

— L'excellence dans l'action, bien entendu ! Cela suppose de savoir à la fois bien juger et bien agir. La science reine en la matière est l'éthique⁷. Le meilleur d'entre vous ne peut être que le plus moral.

D'autres philosophes, attirés par l'agitation, se joignirent à la conversation :

— Débattons ensemble et voyons ce qui ressortira de cette confrontation, mais, en effet, le meilleur selon nous ne peut être que le plus moral !

Un attroupement d'une quarantaine de passants et de sages se fit, tous prêts à intervenir dans le débat si l'opportunité se présentait.

Le robot peut-il être un agent moral ?

La différence ontologique

L'humain, particulièrement hardi, attaqua alors publiquement le robot. Ayant suivi quelques heures de philosophie pour passer l'épreuve du bac, il énonça en vrac tous les éléments habituellement nécessaires pour guider l'action morale dont il put se souvenir : la capacité à souffrir, à avoir des intérêts, des émotions, une conscience, être libre, autonome, responsable, faire des choix..., autant d'éléments qui faisaient défaut au robot selon lui.

— En conséquence, j'accuse le robot de ne pas être une personne morale, de ne pas être conscient, de ne pas avoir de libre arbitre, de ne pas être « incarné », de ne pas comprendre ce qu'il fait et de ne pas même être en vie, contrairement à moi !

Comment situer le robot dans le champ moral ?

— Comment le robot pourrait-il être le meilleur moralement ? L'éthique ne le concerne même pas ! Tu es disqualifié d'office !

Un des philosophes présents, l'Italien Gianmarco Veruggio, intervint :

— Ne le disqualifie pas trop vite. Selon moi, les robots ne sont pas étrangers à l'éthique. En fait, je propose que nous parlions à leur sujet de « roboéthique ».

L'assemblée lui jeta un regard interrogateur. Peter Asaro, spécialiste de la philosophie des sciences, précisa :

— Oui, lorsque nous parlons de l'éthique et des robots, nous pouvons dire trois choses différentes à leur sujet. D'une part, que l'éthique des robots peut être celle de leurs créateurs, qui ont l'obligation morale de ne pas créer de robots dangereux par exemple. Et, d'autre part, ce peut être l'éthique de leurs utilisateurs, qui ne peuvent moralement pas utiliser les robots en leur possession pour commettre de mauvaises actions, par exemple voler quelqu'un, ni maltraiter les robots, car on peut imaginer leur reconnaître des droits un jour.

Le robot, enthousiaste, lança :

— Mais oui, quelle bonne idée !

L'humain coupa court :

— Si ça concerne des droits, ça relève sûrement de la loi, alors attends qu'on croise les avocats pour qu'on en parle. Et, jusqu'ici, je ne vois toujours que l'humain derrière le robot qui soit concerné par les obligations morales...

Asaro poursuivit :

— La troisième possibilité est qu'un robot agisse lui-même en conformité avec l'éthique. Dans ce cas-là, l'éthique le concerne en tant qu'acteur et il reste dans la compétition.

Le précurseur du *computer ethics field*⁸, l'Américain James H. Moor, exposa alors sa propre classification morale des machines.

Au premier degré on trouvait les *ethical impact agents*, « agents ayant un impact éthique », où il regroupait les machines dont on pouvait évaluer les conséquences éthiques. Autrement dit quasiment toutes les machines du monde. Le deuxième degré intégrait les *implicit ethical agents*, « agents moraux implicites ». Il s'agissait de toutes les machines disposant de procédures de sécurité. (Par exemple les machines à couper le pain dans les supermarchés, qui sont dotées d'un gros bouton rouge d'arrêt d'urgence.) L'humain ricana, songeant que tous les ingénieurs ayant respecté leur déontologie de ne pas nuire à l'utilisateur étaient loin de se douter qu'ils créaient là des *implicit ethical agents*. Le troisième intégrait les *explicit ethical agents*, « agents moraux explicites ». Cette classe ne comportait que les machines capables d'un raisonnement éthique grâce à une programmation spécifique. Le quatrième et dernier degré désignait les *full ethical agents*, « agents éthiques "complets" ». Ces robots-là étaient capables d'opérer des jugements moraux et de les justifier, ce qui supposait qu'ils étaient dotés d'une conscience, d'une intentionnalité (la possibilité de penser à quelque chose) et d'un libre arbitre (une liberté de choix). L'humain demanda alors à rencontrer les machines que James H. Moor plaçait dans cette catégorie. Celui-ci, embarrassé, précisa qu'il n'avait pas encore rencontré de modèles robotiques de cette sorte :

— Les ingénieurs devraient se concentrer sur la seule programmation de la morale, parce qu'il ne sera peut-être jamais possible de construire des *full ethical agents*, et sûrement pas dans un avenir proche.

Il quitta la discussion pour partir à la recherche d'autres sages qui pourraient l'aider à construire des *explicit ethical agents*, se promettant de revenir avec des alliés dans ce projet.

Un philosophe recentra la discussion :

— Ne pourrions-nous pas considérer ces machines comme des agents parce qu'elles agissent ? Et, si elles agissent, on peut imaginer évaluer moralement leur action. Le robot serait alors un agent moral comme un autre.

Une levée de boucliers se fit, cette idée semblait heurter certains anciens. Cela ne venait à l'esprit de personne de considérer le Soleil, la force de gravitation ou même les médicaments comme des agents, sous prétexte qu'ils provoquaient des effets ! La chose semblait entendue : provoquer un effet n'était pas suffisant pour être un agent moral.

Asaro en profita pour reprendre la parole :

— *« C'est une tentation de penser qu'il n'y a que deux types distincts d'agents causaux⁹ dans le monde : les agents amoraux¹⁰ et les agents moraux. Au lieu de cela, je suggère qu'il serait utile de penser l'agence morale comme un continuum allant de l'amoralité à la pleine moralité autonome¹¹. »*

Même si l'agence morale était un continuum dont il n'était pas directement exclu, jusqu'ici le robot semblait néanmoins en mauvaise posture. Quels que soient la classification retenue et le niveau auquel il pouvait prétendre, il semblait toujours loin derrière les capacités morales de l'humain. Ce dernier se réjouissait déjà de sa première victoire en se frottant les mains.

Le robot se défendit d'un cri déchirant :

— Pourquoi me refusez-vous le droit d'être un agent moral ? Vous ne faites que produire une nouvelle forme de racisme ! Vous méprisez les robots comme naguère le colonialiste méprisait les ethnies extra-occidentales !

Jeremy Bentham intervint :

— Du calme... Plutôt que du racisme, c'est peut-être une nouvelle forme de spécisme: nous défendons notre espèce plutôt que ceux de ta sorte. J'avais de mon temps défendu les droits des animaux avec la formule: «*The question is not, Can they reason? nor, Can they talk? but, Can they suffer*¹²? » Malheureusement, toi, tu n'as pas de système nerveux pour souffrir.

— Non, répondit le robot, mais cela n'a pas d'importance. Après tout, vous vous occupez bien des patients en coma végétatif, non? Je suis peut-être moi aussi juste un patient moral dont vous devez prendre soin. Si vous trouvez que ces patients-là, qui ne sentent plus rien et n'ont conscience de rien non plus, ont droit à votre considération morale, alors mon incapacité à souffrir ou à avoir conscience de moi-même n'est pas un critère pour me refuser des droits moraux!

Un éthicien environnemental et juriste, Christopher D. Stone, renchérit:

— Moi, j'ai défendu le droit des plantes! Dans mon livre de 1972, *Should Trees Have Standing?*, je soutiens qu'un être non conscient n'en a pas moins des intérêts dans la vie, par exemple grandir, et qu'il serait cruel de l'en priver, en le tronçonnant sauvagement. Les organismes ont donc des bio-intérêts que nous devons respecter comme un droit moral.

John Basl chuchota à Stone:

— Là, je ne suis pas sûr que tu aides notre ami en fer-blanc. Tu n'as peut-être pas remarqué, mais ce n'est pas un organisme biologique, il n'est pas vivant. Enfin... jusqu'ici du moins! Une version plus subtile de ton argument serait de ne plus parler de bio-intérêts, mais de téléo-intérêts. Parce que ce que tu veux

protéger moralement, c'est la poursuite du *télos* (d'un «but»), comme grandir tu l'as dit, et pas le fait d'appartenir à une classification biologique. Peu importe qu'il soit vivant ou pas.

Christopher D. Stone l'interrogea :

— Mais qu'est-ce que ça change pour lui qu'on parle de téléo-intérêts ?

— À la fois beaucoup, et rien au fond, répondit Basl.

— Mais pourquoi ? demanda le robot. Je suis bien un être non conscient qui poursuit un but, un *télos*, grâce à ma programmation. J'ai des cousins aspirateurs dont le seul intérêt dans la vie est de passer partout dans une pièce et qui font tout pour y arriver, même s'ils sont gênés sur leur chemin.

— Justement, soupira Basl, vos buts proviennent de votre programmation. Cela veut dire que vos téléo-intérêts ne sont rien d'autre que l'usage que nous souhaitons avoir de vous-mêmes. Tu crois que tes cousins suivent leur volonté en aspirant consciencieusement ? Pas du tout, c'est la nôtre qu'ils suivent ! Si je bloque un aspirateur avec mon pied parce qu'il allait faire tomber une lampe, je ne nuis pas à la poursuite de son but, parce qu'il n'a pas été conçu pour casser des bibelots. Et, si je le bloque parce que ça m'amuse, mon propre intérêt psychologique de recherche d'amusement l'emporte sur celui de tes cousins dont j'ai l'usage. Autrement dit, téléo-intérêt ou pas, les robots n'ont pas de droits moraux que je suis tenu de respecter. D'ailleurs, la seule chose qui m'empêche de te mettre en pièces là tout de suite, c'est que tu appartiens sûrement à quelqu'un et que je respecte son droit de propriété. Mais je te prie de croire que si tu étais à moi j'aurais parfaitement le droit de t'arracher les antennes si ça me chantait.

Le robot sembla scandalisé.