

Bayerische Julius-Maximilians-Universität Würzburg

Institut Neuphilologie

Proseminar: "Littérature américaine de science-fiction"

MA Matthias Kemmer

Semestre d'été 2010

Développements des « Trois lois de la robotique » et

caractéristiques de chaque étape :

Une discussion basée sur la série Robot d'Asimov

16.08.2010

Xin Chen

Haugerglasisstr.8

97080 Würzburg

Anglistik/Amerikanistik, Études politiques et sociales (BA)

2ème semestre

Matr. N° 1683950

Développements des « Trois lois de la robotique » :

Une discussion basée sur la série Robot d'Asimov

Table des matières

1. Remarque mentale	... 3
2. Introduction : La tendance au développement des Trois Lois... 4	
3. Première étape : L'obéissance absolue aux trois lois	... 5
3.1 Interprétation des règles	... 6
3.2 Contradictions et « robot-blocage »	... 7
3.3 Difficultés à évaluer les situations et à prendre des décisions... 8	
4. Deuxième étape : La capacité de jugement et son impact sur les lois... 9	
3.1 Impact sur la première loi	... 9
3.2 Impact sur la deuxième loi	... dix
5. Troisième étape : Les trois lois étendues de la robotique	... 11
5.1 La création de la loi zéro	... 11
5.2 La conséquence de la création de la loi zéro	... 12
6. Référence	... 14

Développements des « Trois lois de la robotique » :

Une discussion basée sur la série Robot d'Asimov

2. Introduction

L'idée de rechercher les trois lois de la robotique est venue d'une question – Une question importante dont nous avons discuté à plusieurs reprises dans le séminaire : « Quelle est la différence entre les humains et les robots ? »

Avec l'évolution des technologies modernes et l'émergence de cyborgs, définir la frontière entre les humains et les robots devient de plus en plus difficile. Avec des gens équipés d'organes artificiels, on ne peut plus argumenter du point de vue de la physique ou de la biologie. La notion d'humain les êtres humains commence à se confondre avec celui des robots. Pendant ce temps, l'intelligence artificielle acquiert également progressivement des caractéristiques humaines grâce à leur capacité d'apprentissage.

Alors, quelle est la différence entre les êtres humains et les robots ?

Une des réponses serait les « trois lois de la robotique » qui empêcher les robots d'agir arbitrairement ou aussi librement que les êtres humains.

1. Un robot ne peut pas blesser un être humain ni, par inaction, permettre un être humain à subir un préjudice.
2. Un robot doit obéir à tous les ordres qui lui sont donnés par des êtres humains, sauf lorsque de tels ordres seraient en conflit avec la Première Loi.
3. Un robot doit protéger sa propre existence aussi longtemps que celle-ci la protection n'est pas incompatible avec la première ou la deuxième loi.

– Moi, Robot d'Isaac Asimov

Ce document de recherche traite des lois qui ont été créées par Isaac Asimov – Sans doute le plus grand écrivain de science-fiction de tous les temps.

Dans le monde fictif d'Asimov, tous les robots sont soumis à une interprétation stricte de ces règles. L'obéissance absolue aux trois lois résulte la plus distinction remarquable entre les humains et les robots.

Cependant, les trois lois de la robotique se développent tout au long de la vie. Série de robots. Au cours du développement, nous pouvons sentir le clair intention de brouiller la frontière entre les robots et les humains. Aussi, plus

Développements des « Trois lois de la robotique » :

Une discussion basée sur la série Robot d'Asimov

plus important encore, en étudiant l'évolution des trois lois, nous pouvons également découvrir comment évolue la relation entre la robotique et l'être humain.

Pour avoir une image claire de la tendance du développement, je
a organisé l'information en trois étapes.¹

Dans une première étape : j'ai analysé les cas dans lesquels les robots obéissent aux trois lois absolument. Dans cette étape, liés par les trois lois, les robots sont subordonnés aux êtres humains. Les robots étant incapables de juger, la différence entre les humains et les robots est remarquable.

Dans la deuxième étape : les robots ont reçu la capacité de réaliser jugements comme les êtres humains. Les trois lois sont assouplies, ce qui a produit des robots dotés d'une autonomie considérable. A ce stade, il est impossible de faire la différence entre un robot et le meilleur des humains.

Enfin, je présenterai la loi zéro, son impact sur la coexistence des robots et des humains et des problèmes nouvellement apparus.

3. Première étape : L'obéissance absolue aux trois Lois

Dans le monde fictif d'Asimov, les trois lois de la robotique sont construites le plus profondément dans le cerveau positronique d'un robot en tant que code moral immuable pour eux. L'existence des robots repose sur l'obéissance absolue de ces derniers.

trois règles. Tout signe d'infraction à la loi détruirait le robot ou le mettrait dans une impasse. Pour prouver ce point, un exemple peut être trouvé dans le histoire « Rêves de robots ». Le robot Elvex avait rêvé d'ignorer les lois et libérant les robots de l'esclavage humain et fut donc démantelé par le Dr. Susan Calvin par souci de sécurité.

Le but des Trois Lois d'Asimov est de montrer clairement que les robots et Les machines en général ne sont que des constructions artificielles et peuvent être entièrement contrôlées². Toutefois, l'application des lois est problématique, car

¹ Remarque : Les trois étapes ne sont pas classées dans un ordre chronologique, ni les histoires utilisées pour les illustrer.

² Portelli, Alessandro. "Trois lois de la robotique, Les : lois du texte, lois de la production, lois de la société." Études de science-fiction 7(2) :150-156. Juillet 1980.

Développements des « Trois lois de la robotique » :

Une discussion basée sur la série Robot d'Asimov

les trois lois sont macroscopiques, elles ne peuvent pas être appliquées à chaque situation sans jugement indépendant.

Dans la première partie de l'article, j'analyse les caractéristiques et problèmes rencontrés par les robots à ce stade, qui peuvent également être considérés comme le différences significatives entre les robots et les humains.

3.1. Interprétation des règles

Premièrement, l'interprétation de la langue des commandes peut être problématique pour les robots, car ils pourraient ne pas être capables de comprendre expressions figuratives. Dans l'histoire mentionnée ci-dessus³, le Dr Calvin met robot Elvex en mode veille en disant « Elvex, tu ne bougeras ni ne parleras et ne nous écoute pas jusqu'à ce que je prononce à nouveau ton nom. Obéissant à la deuxième loi, le robot semblait être devenu une seule pièce de métal. En interprétant cette commande, un l'humain assisterait probablement encore à une fonction corporelle passive. Cependant, les robots se limitent à une pure interprétation littérale, et feront absolument l'affaire rien sur le sujet concerné jusqu'à ce que la commande suivante se produise. Contrairement à êtres humains, les robots sont beaucoup moins susceptibles de reconnaître quand il y a un changement de circonstances qui rend impossible le prochain commandement arriver. Ainsi, ils resteraient là même dans une situation très urgente.

En réalité, il est naïf de penser qu'il existe un sens littéral à chaque commande. Les robots doivent donc faire preuve de jugement pour interpréter les situation, puis décidez comment réagir.

La première loi fondamentale de la robotique stipule : « un robot ne peut pas blesser un être humain ou, par inaction, lui permettre de subir un danger. Dans ici, préjudice désigne TOUT type de blessure. Non seulement physique, mais aussi psychologique.

3 Asimov, Isaac. «Rêves de robots». Rêves de robots. New York : Ace Books, 1987. P.39

Développements des « Trois lois de la robotique » :

Une discussion basée sur la série Robot d'Asimov

Toujours dans « Petit robot perdu », Asimov décrit la situation : Parce que de la première loi stricte, les robots étaient surprotecteurs et donc rendus le travail en champ gamma impossible.⁴

Comparativement aux blessures physiques, les dommages psychologiques créent davantage de complications, car l'esprit humain est difficile à comprendre pour les robots. Dans l'histoire « Menteur », un robot qui lit dans les pensées a déclaré : « vous n'avez aucune idée à quel point c'est compliqué ils (l'esprit humain) le sont. Je n'arrive pas à tout comprendre parce que mon propre esprit a si peu de choses en commun avec eux. »⁵ Dans une partie ultérieure de l'histoire, Lorsque le robot a été confronté aux questions des humains, il a choisi de leur mentir afin d'éviter de les blesser directement. Cependant, dans la fin, finalement, en découvrant la vérité, les gens ont été encore plus blessés. C'est le résultat de la fausse interprétation des émotions humaines par le robot. Les êtres humains pourraient comprendre qu'un mensonge brisé peut causer plus de dégâts qu'une vérité désagréable. En raison de la mauvaise compréhension de l'esprit humain, le robot n'a pas réussi à sentir les conséquences du mensonge.

3.2. Contradictions et blocage des robots

Un autre problème des trois lois survient lorsqu'il y a une contradiction. Lorsque le robot reçoit une commande qui entre en conflit avec une commande intégrée. En règle générale, il souffrira d'un blocage robot qui l'empêchera de fonctionner.

Ce problème a été présenté dans la toute première histoire dans laquelle Asimov a officiellement introduit les trois lois. Dans l'histoire, le robot est entré dans un gel mental (blocage du robot) car il ne pouvait pas décider s'il devait effectuer la tâche de retrouver l'équilibre qui pourrait le mettre en danger. Habituellement, le robot n'a le droit de protéger sa propre existence que lorsqu'il n'est pas en contradiction avec la première et la deuxième loi. Cependant, dans ce cas, la commande était très faible et le robot était programmé pour avoir une personnalité très forte. tendance à la protection, de sorte que le potentiel de la troisième loi est inhabituellement élevé. Donc,

⁴ Asimov, Isaac. "Petit robot perdu." JE ROBOT. New York : Bantam Books, 2004. P.117

⁵ Asimov, Isaac. "Repaire." JE ROBOT. New York : Bantam Books, 2004. P.95

Développements des « Trois lois de la robotique » :

Une discussion basée sur la série Robot d'Asimov

les conflits entre la deuxième loi faible et la troisième loi forte étaient établis et abouti à ce que le robot suive « [...] un cercle autour, en restant sur le lieu de tous les points d'équilibre potentiel. »⁶

De plus, un Robot-bloc est également possible au sein d'une seule loi. Encore une fois dans l'histoire "Lair", le Dr Calvin a indirectement "tué" le robot lisant les pensées Herbie en le confrontant à un dilemme insoluble. Après avoir répété : « Vous ne pouvez pas dis-leur, parce que ça ferait mal et tu ne dois pas faire mal, mais si tu ne le dis pas eux, tu as mal, alors tu dois leur dire... »⁷ le robot « ...est mort dans néant, et s'est effondré en un tas d'immobilités.

3.3. Difficultés d'évaluation et de prise de décision

Dans la première loi : un robot ne doit nuire à AUCUN être humain ni laisser AUCUN l'être humain est blessé. Si la première loi est absolue, il ne devrait y avoir aucune valorisation de tout individu humain. Donc tout le monde est complètement égal dans l'évaluation du robot. Des problèmes peuvent facilement être soulevés ici. Par exemple, un voleur attaque un innocent. En pensant que la première loi n'exige pas de robot pour juger entre une « bonne » ou une « mauvaise » personne, il est probable que le robot aiderait les deux côtés. Dans une situation d'urgence, cela peut même n'aider que celui avec une meilleure chance de survivre, même si celle-ci pourrait être une criminel.

Des difficultés plus graves surviennent en ce qui concerne la valorisation des titres multiples humains. Hypothétiquement, dans cette condition, toute action entreprise par le robot cela entraînerait certainement un préjudice pour l'être humain. Ensuite le robot doit sélectionner quelle action entraîne le moins de préjudice. Cependant, les robots ne peuvent pas juger comme les êtres humains le font, car « il n'est pas pratique pour les robots de fabriquer jugements sur des points fins tels que le talent, l'intelligence, l'utilité générale à la société car cela retarderait la décision au point où le robot

⁶ Asimov, Isaac. "Courir autour." JE ROBOT. New York : Bantam Books, 2004. P.38

⁷ Asimov, Isaac. "Repaire." JE ROBOT. New York : Bantam Books, 2004. P.110

Développements des « Trois lois de la robotique » :

Une discussion basée sur la série Robot d'Asimov

est effectivement immobilisé. Alors ils procèdent par chiffres. »⁸ En conséquence, lorsqu'un Le robot doit choisir entre sauver un scientifique précieux ou cinq criminels, il choisirai certainement plus tard.

4. Deuxième étape : la capacité de jugement et son impact sur les lois

Dans cette étape, j'écrirai sur les cas dans lesquels le jugement la capacité a été insérée dans les robots et l'impact que la capacité a sur les trois lois.

Dans une histoire ultérieure « Tu es attentif à lui », après que les robots aient été utilisé sur les stations spatiales et les colonies planétaires depuis des décennies, les États-Unis La tentative de Robot de mettre des robots sur terre. Considérant les problèmes dans la première étape, ils ont décidé d'insérer la capacité de jugement humain dans les robots cerveau positronique. Pouvant juger comme un humain, les robots ne sont plus dans le cadre strict des deux premières lois. Ils ont acquis leur autonomie dans un certain degré.

4.1. L'impact sur la Première Loi

Dans l'histoire, étant donné la capacité de juger, les robots «...peuvent peser chaque l'être humain en termes de sexe, d'âge, de position sociale et professionnelle, intelligence, maturité, responsabilité sociale, etc.

Sous cette condition ci-dessus, la Première Loi sera affectée lorsque chacune des actions de protection possibles choisies par un robot nuira au personnes. Le robot doit alors agir selon son jugement. Différent dès la première étape, non seulement la quantité mais aussi la qualité de l'humain les êtres seraient considérés ici. En cas d'urgence, les robots choisir d'assurer la sécurité de l'être humain le plus apte mentalement, caractère et connaissances.

⁸ Asimov, Isaac. «Tu te souviens de lui.» Le robot complet. Londres : Grenade, 1982. P.498

Développements des « Trois lois de la robotique » :

Une discussion basée sur la série Robot d'Asimov

Les individus ne sont donc plus égaux face aux robots. Ils ont appris comment construire une hiérarchie entre les gens et ils ont commencé à choisir lequel devrait être sacrifié pour le bien des plus capables.

4.2. L'impact sur la deuxième loi

Dans la première étape, la Deuxième Loi exige que les robots obéissent à tous. ordres humains. Pourtant, les robots sont confrontés au problème : sur Terre, ils viendraient auprès d'une grande variété de personnes, qui ne sont pas toutes fiables ou responsables.

En conséquence, la Deuxième Loi a été profondément modifiée. Chaque action potentielle doit impliquer le propre jugement du robot. Ils obéissent davantage rationnellement, par rapport à l'obéissance absolue de la première étape.

Premièrement, la définition des êtres humains dans la deuxième loi a été rétréci. Les robots n'obéissent plus à TOUS les humains, mais seulement à un seul. qui est apte par son esprit, son caractère et ses connaissances à donner des ordres.

Deuxièmement, les robots commencent à apprendre à appliquer la deuxième loi. correctement, les contradictions et les conflits entre les ordres ne sont plus problématique. Ils peuvent décider à quel être humain obéir et lequel ne pas obéir. obéissez en fonction de leur maturité, de leur intelligence et de leur niveau d'éducation.

En donnant aux robots la capacité de juger, Asimov a brouillé les frontière entre les êtres humains et les robots.

Même dans une des premières histoires « Evidence », Asimov a créé un mystérieux personnage – Stephen Byerley. Tout au long de l'histoire, Asimov a essayé de laisser entendre que Byerley était en fait un humanoïde.

Une fois en classe, en argumentant d'un point de vue intelligent, j'ai essayé de prouver l'identité humaine de Byerley : Selon la formulation stricte de la première loi, les robots ne doivent nuire à aucun être humain, non seulement physiquement, mais également psychologiquement. Les sentiments affectifs de déception peuvent être considérés comme préjudice psychologique. En insistant sur le fait d'être humain, Byerley a déçu son rival politique Quinn et a ensuite enfreint la première loi de la robotique. On devrait donc pouvoir dire qu'il n'est pas un robot.

Développements des « Trois lois de la robotique » :

Une discussion basée sur la série Robot d'Asimov

Cependant, sous la condition de la deuxième étape, mon argument « intelligent » est intenable. Parce que les robots avancés sont capables de juger quel est le meilleure action à entreprendre pour le bien des êtres humains. Chez Stephen Byerley Dans ce cas, il tenterait de remporter les élections afin d'aider et de protéger davantage personnes. Même s'il doit d'abord battre Quinn aux élections et le décevoir un peu.

Malgré toutes les allusions données par Asimov dans l'histoire, les lecteurs ne peuvent pas dire pour autant sûr qu'il s'agisse d'un être humain ou d'un robot humanoïde. C'est car, lorsqu'un robot est doté de la capacité de jugement, « le trois règles de la robotique sont (aussi) les principes directeurs essentiels d'une bonne de nombreux systèmes éthiques dans le monde... Vous ne pouvez tout simplement pas faire la différence entre un robot et le meilleur des humains. 9

5. Troisième étape : les trois lois étendues de la robotique

Dans les années 50, Asimov montrait déjà des signes d' extension de la première loi de protéger les humains individuels pour protéger l'humanité entière.

Dans l'histoire « The Evitable Conflict », les puissants ordinateurs positroniques ont commencé à aller à l'encontre de leurs fonctions. Alors que les robots positroniques sont devenus plus développés, ils ont progressé au-delà de la possibilité d'une étude détaillée contrôle humain. Après enquête, les gens réalisent enfin que le les dysfonctionnements sont les actes délibérés des Machines. Parce que l'ordinateur « avoir à cœur le bien de l'humanité grâce à la force écrasante du Première loi de la robotique »¹⁰, ils ont eux-mêmes dégénéralisé la première loi ce qui leur a permis de laisser un petit préjudice arriver à certains individus afin de protéger une grande quantité de dommages causés à l'humanité comme un ensemble.

9 Asimov, Isaac. "Preuve." JE ROBOT. New York : Bantam Books, 2004. P.182

10 Asimov, Isaac. « Le conflit inévitable ». JE ROBOT. New York : Bantam Books, 2004. P.201

Développements des « Trois lois de la robotique » :

Une discussion basée sur la série Robot d'Asimov

5.1. La création de la loi zéro

Cette Première Loi dégénéralisée est très similaire à la Loi Zéro, qui Asimov a formulé en 1985 dans Robot and Empire :

Avant la création de la loi Zeroth, un robot devait protéger une personne même si la survie de l'humanité dans son ensemble était mise en danger. Même dans la deuxième étape, la capacité de jugement des robots était limitée entre individus humains.

Dans l'histoire ci - dessus, les deux robots – Daneel et Giskard rassembler les preuves d'un plan génocidaire massif élaboré par Amadiro, en utilisant intensificateur nucléaire, un roboticien spatial veut rendre la Terre inhabitable en êtres humains et héritent donc de la Galaxie Lactée. Les robots essaient de s'arrêter le plan, mais ils sont gênés par la première loi de la robotique qui empêche les empêcher de nuire directement à Amadiro.

Finalement, face à cette situation urgente, le robot Daneel a conclu le Zeroth Loi:

« L'humanité dans son ensemble est plus importante qu'un seul être humain.

Il existe une loi qui est plus grande que la Première Loi : un robot ne peut pas blesser l'humanité, ou par l'inaction, permettre à l'humanité de subir des dommages. »

- Robot et Empire par Isaac Asimov

5.2. La conséquence de la création de la loi zéro

La loi Zeroth crée une complicité accrue dans la prise de décision, parce que le mot humanité n'est qu'un concept abstrait. C'est difficile de définir l'humanité et, par conséquent, difficile de décider de ce qui est bon pour elle. Comme mentionné dans la deuxième étape, il est déjà assez difficile de décider quel individu peut souffrir, plus le préjudice est grand. Il est encore plus difficile de choisir entre un l'individu et l'humanité, quand on ne sait même pas exactement quel aspect de l'humanité à laquelle il a peut-être affaire. Estimation des probabilités par rapport à la gestion de groupes d'humains est bien plus difficile qu'avec des humains individuels.

Développements des « Trois lois de la robotique » :

Une discussion basée sur la série Robot d'Asimov

Outre les problèmes ci-dessus, la création de la loi Zeroth a également de graves conséquences concernant la relation entre les robots et humains.

Ayant la responsabilité de protéger l'humanité entière, les robots étaient plus susceptibles d'adopter une attitude paternaliste envers les humains. Avec le les robots se considèrent comme des êtres plus intelligents et plus capables, les La capacité de l'humanité à contrôler son propre destin est compromise.

Dans « The Evitable Conflict », les bases de la domination des robots ont été déjà posé : « La Machine mène notre avenir à notre place... L'humanité a perdu son mot à dire sur son avenir. ¹¹ Cette idée a été profondément développée dans Le dernier roman robotique d'Asimov – Robot and Empire. Différent de tous les autres précédents Dans les histoires écrites, les protagonistes ne sont pas des êtres humains, mais deux robots. Ils Il a conclu : « Il ne suffit donc pas de choisir, ami Daneel. Nous devons être capable de façonner, de façonner une espèce désirable puis de la protéger.¹² Exemples des robots règnent sur les humains se trouve dans la dernière partie du roman, avec la permission de la Loi Zéro, ils ont manipulé les humains pour qu'ils agissent de manière une manière qui, selon eux, résoudrait les problèmes. Pendant ce temps, ils ont également fait un décision importante pour l'humanité – abandonner la terre, en permettant au création d'une terre radioactive.

La loi Zeroth est le symbole de la supériorité des robots sur humanité. Comme il s'agit de la dernière étape de l'élaboration des Lois, le la transition de l'esclavage des robots vers la domination des robots est terminée. Quant à la question : « Quelle est la différence entre les êtres humains et les robots ? J'espère que le le papier répond à certaines.

Cependant, si la vérité reste encore vague, je citerai ce que dit le Dr Calvin a dit un jour : « Vous n'avez pas travaillé avec eux, donc vous ne les connaissez pas. Ils sont une meilleure race que nous... »¹³

¹¹ Asimov, Isaac. « Le conflit inévitable ». JE ROBOT. New York : Bantam Books, 2004. P.224

¹² Asimov, Isaac. Robot et Empire. Londres : HarperCollins, 1996.

¹³ Asimov, Isaac. JE ROBOT. New York : Bantam Books, 2004. PXI

Développements des « Trois lois de la robotique » :

Une discussion basée sur la série Robot d'Asimov

6. Bibliographie

1. Asimov, Isaac. "Preuve." JE ROBOT. New York : Livres Bantam, 2004.
2. Asimov, Isaac. "Repaire." JE ROBOT. New York : Livres Bantam, 2004.
3. Asimov, Isaac. "Petit robot perdu." JE ROBOT. New York : Livres Bantam, 2004.
4. Asimov, Isaac. "Courir autour." JE ROBOT. New York : Livres Bantam, 2004.
5. Asimov, Isaac. «Rêves de robots». Rêves de robots. New York : Ace Books, 1987.
6. Asimov, Isaac. Robot et Empire. Londres : HarperCollins, 1996.
7. Asimov, Isaac. « Le conflit inévitable ». JE ROBOT. New York : bantam Livres, 2004.
8. Asimov, Isaac. «Tu te souviens de lui.» Le robot complet. Londres : Grenade, 1982.
9. Portelli, Alessandro. « Trois lois de la robotique, Les : lois du texte, lois de la production, les lois de la société. Études de science-fiction 7(2) :150-156. Juillet 1980.